

**PERINGKASAN TEKS MULTI DOKUMEN BERBAHASA INDONESIA  
MENGUNAKAN *SENTENCE SCORING* DAN  
*SUPPORT VECTOR MACHINE (SVM)***

**SKRIPSI**

**Oleh:  
DERI FAUZI  
NIM. 19650074**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**PERINGKASAN TEKS MULTI DOKUMEN BERBAHASA INDONESIA  
MENGUNAKAN *SENTENCE SCORING* DAN *SUPPORT VECTOR  
MACHINE (SVM)***

**SKRIPSI**

Diajukan Kepada :  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk Memenuhi Salah Satu Persyaratan Dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)

**Oleh:  
DERI FAUZI  
NIM. 19650074**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**HALAMAN PERSETUJUAN**

**PERINGKASAN TEKS MULTI DOKUMEN BERBAHASA INDONESIA  
MENGUNAKAN *SENTENCE SCORING* DAN *SUPPORT VECTOR  
MACHINE (SVM)***

**SKRIPSI**

**Oleh:  
DERI FAUZI  
NIM. 19650074**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 28 November 2023

Pembimbing I,



Dr. Zainal Abidin, M.Kom  
NIP. 19760613 200501 1 004

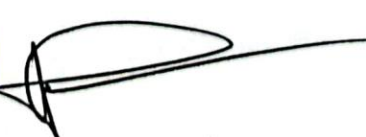
Pembimbing II,



Fatchurrochman, M.Kom  
NIP. 19700731 200501 1 002

Mengetahui,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Fachrul Kurniawan, ST., M. MT., IPM  
NIP. 19771020 200912 1 001

## HALAMAN PENGESAHAN

### PERINGKASAN TEKS MULTI DOKUMEN BERBAHASA INDONESIA MENGUNAKAN *SENTENCE SCORING* DAN *SUPPORT VECTOR MACHINE (SVM)*

#### SKRIPSI

Oleh:  
**DERI FAUZI**  
**NIM. 19650074**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Komputer (S.Kom)  
Pada Tanggal: 11 Desember 2023

#### Susunan Dewan Penguji

Ketua Penguji : A'la Syauqi, M.Kom  
NIP. 19771201 200801 1 007

Anggota Penguji I : Shoffin Nahwa Utama, M.T  
NIP. 19860703 202012 1 003

Anggota Penguji II : Dr. Zainal Abidin, M.Kom  
NIP. 19760613 200501 1 004

Anggota Penguji III : Fatchurrochman, M.Kom  
NIP. 19700731 200501 1 002

(  )


(  )

(  )

(  )

Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Fachrul Kurniawan, ST., M. MT., IPM  
NIP. 19771020 200912 1 001

## PERNYATAAN KEASLIAN PENULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Deri Fauzi

NIM : 19650074

Fakultas / Prodi : Sains dan Teknologi / Teknik Informatika

Judul Skripsi : Peringkasan Teks Multi Dokumen Berbahasa Indonesia  
Menggunakan *Sentence Scoring* dan *Support Vector Machine (SVM)*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 11 Desember 2023

Yang membuat pernyataan,



Deri Fauzi

NIM.19650074

## HALAMAN MOTTO

*“Tidak ada ujian yang tidak bisa diselesaikan. Tidak ada kesulitan yang melebihi batas kesanggupan. Karena Allah swt. Tidak akan membebani seseorang melainkan sesuai dengan kadar kesanggupannya”*

*(Q.S Al-Baqarah: 286)*

*“Setelah diterpa oleh hujan siapapun pasti bisa jadi semakin kuat. Segelap apapun harimu, dari tengah awan hujan pasti ada langit penuh warna”*

*(JKT48 – After Rain)*

## **HALAMAN PERSEMBAHAN**

Penulis ingin mempersembahkan skripsi ini kepada orang tua, keluarga, dosen, sahabat, dan semua pihak yang telah membantu secara aktif dalam menyelesaikan penelitian ini

## KATA PENGANTAR

*Assalamu'alaikum Wr.Wb.*

Puji syukur penulis panjatkan kepada Allah SWT yang senantiasa memberikan rahmat dan kesehatan, sehingga penulis mampu menyelesaikan skripsi ini dengan baik. Penulis menyampaikan ucapan Terimakasih kepada semua pihak yang pernah terlibat langsung maupun tidak langsung dalam menyelesaikan skripsi ini, bukan hanya karena usaha keras dari penulis sendiri, akan tetapi karena adanya dukungan dari berbagai pihak. Oleh karena itu penulis berterima kasih kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, S.T., M.MT., IPM selaku Ketua Prodi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Zainal Abidin, M.Kom selaku Dosen Pembimbing 1 dan Fatchurrochman, M.Kom selaku Dosen Pembimbing 2 yang telah membimbing serta memberikan arahan serta motivasi dalam penulisan skripsi dari awal hingga akhir.
5. Para dosen dan staf akademika Jurusan Teknik Informatika yang telah memberikan ilmu yang sangat bermanfaat, yang daripada itu secara tidak langsung telah turut serta dalam penyusunan skripsi ini
6. Kedua orang tua penulis, Bapak Moh. Jazuli, Ibu Enik Winarsih dan kakak saya Latief Zulfikar Chusaini yang telah memberikan banyak dukungan, doa serta bantuan sehingga penulis mampu menyelesaikan masa studi hingga mencapai gelar sarjana.



7. Keluarga besar Program Studi Teknik Informatika terutama Angkatan 2019 ALIEN “*Alliance of Informatics Engineering*” yang telah memberikan dukungan untuk saling menyelesaikan skripsi.
8. Saudara seperjuangan saya yang terkumpul dalam paguyuban “Ahlussunnah Wal Jama’ah” Bisyri, Thoriq, Fikri, Sadad, Zulfan, Sadad, Dicky, Riduan, Pejon, Andi, Alfin, Anam, Dayat, Faiz Farros, Faiz Hilmi, Puput, Widia, dan Bila serta satu orang spesial bagi saya Nur Azizah Munir yang senantiasa memberikan ide, solusi, dan motivasi dan tak luput memberikan kritik, saran, dan informasi dalam masa perkuliahan sampai proses penyusunan skripsi.
9. Penulis juga ingin mengucapkan terima kasih yang sebesar-besarnya kepada JKT48 melalui lagu-lagu mereka. Musik dan lirik dari karya-karya JKT48 telah memberikan inspirasi dan semangat tambahan dalam menyelesaikan penelitian ini.

Penulis sadar bahwa skripsi ini masih sangat jauh dari kata sempurna dan mungkin terdapat kesalahan di dalamnya. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk mengembangkan skripsi ini agar lebih bermanfaat bagi dirinya dan pembaca pada umumnya.

Malang, 11 Desember 2023

Penulis

## DAFTAR ISI

<b>HALAMAN PENGAJUAN</b> .....	<b>ii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iii</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>iv</b>
<b>PERNYATAAN KEASLIAN PENULISAN</b> .....	<b>v</b>
<b>HALAMAN MOTTO</b> .....	<b>vi</b>
<b>HALAMAN PERSEMBAHAN</b> .....	<b>vii</b>
<b>KATA PENGANTAR</b> .....	<b>viii</b>
<b>DAFTAR ISI</b> .....	<b>x</b>
<b>DAFTAR GAMBAR</b> .....	<b>xii</b>
<b>DAFTAR TABEL</b> .....	<b>xiii</b>
<b>ABSTRAK</b> .....	<b>xiv</b>
<b>ABSTRACT</b> .....	<b>xv</b>
مستخلص البحث.....	<b>xvi</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Pernyataan Masalah.....	5
1.3 Tujuan Penelitian.....	5
1.4 Batasan Masalah.....	5
1.5 Manfaat Penelitian.....	6
<b>BAB II STUDI PUSTAKA</b> .....	<b>7</b>
2.1 <i>Text Summarization</i> .....	7
2.2 <i>Text Preprocessing</i> .....	9
2.3 <i>Sentence Scoring</i> .....	10
2.4 <i>Support Vector Machine</i> .....	11
2.5 ROUGE .....	13
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>14</b>
3.1 Pengumpulan Data .....	14
3.2 Desain Sistem .....	15
3.3 <i>Preprocessing</i> .....	16
3.3.1 <i>Case Folding</i> .....	17
3.3.2 <i>Remove Punctuation and Special Character</i> .....	17
3.3.3 <i>Stopword Removal</i> .....	18
3.3.4 <i>Stemming</i> .....	18
3.4 Implementasi <i>Sentence Scoring</i> .....	19

3.4.1 Panjang Kalimat (F1).....	19
3.4.2 Posisi Kalimat (F2).....	21
3.4.3 Kalimat Mengandung Data Numerik (F3).....	22
3.4.4 <i>TF-ISF</i> (F4).....	22
3.4.5 Kemiripan Antar Kalimat (F5) .....	24
3.5 Implementasi <i>Support Vector Machine</i> .....	25
3.6 Proses Peringkasan .....	30
<b>BAB IV UJI COBA DAN PEMBAHASAN .....</b>	<b>31</b>
4.1 Dataset.....	31
4.2 Skenario Uji Coba .....	34
4.3 Hasil Uji Coba.....	35
4.3.1 Percobaan <i>Fold 1</i> .....	36
4.3.2 Percobaan <i>Fold 2</i> .....	38
4.3.3 Percobaan <i>Fold 3</i> .....	40
4.3.4 Percobaan <i>Fold 4</i> .....	42
4.3.5 Percobaan <i>Fold 5</i> .....	43
4.4 Pembahasan.....	45
<b>BAB V PENUTUP.....</b>	<b>49</b>
5.1 Kesimpulan.....	49
5.2 Saran.....	50
<b>DAFTAR PUSTAKA</b>	
<b>LAMPIRAN</b>	

## DAFTAR GAMBAR

Gambar 3. 1 Data dari INDOSUM .....	14
Gambar 3. 2 Desain Sistem.....	15
Gambar 3. 3 <i>Flowchart</i> Praprocessing.....	16
Gambar 3. 4 contoh berita yang akan di <i>preprocessing</i> .....	20
Gambar 3. 5 <i>Flowchart Support Vector Machine</i> .....	26
Gambar 3. 6 <i>Flowchart hyperplane</i> .....	28
Gambar 4. 1. Dataset Story Multi Dokumen.....	32
Gambar 4. 2. <i>K-Fold Cross Validation</i> .....	35
Gambar 4. 3 Hasil Ringkasan Manusia Story 8 .....	37
Gambar 4. 4 Hasil Ringkasan Sistem Story 8 .....	38
Gambar 4. 5 Hasil Ringkasan Manusia Story 11 .....	39
Gambar 4. 6 Hasil Ringkasan Sistem Story 11 .....	39
Gambar 4. 7 Hasil Ringkasan Story 27 .....	41
Gambar 4. 8 Hasil Ringkasan Sistem Story 27 .....	41
Gambar 4. 9 Hasil Ringkasan Manusia Story 34 .....	43
Gambar 4. 10 Hasil Ringkasan Sistem Story 34 .....	43
Gambar 4. 11 Hasil Ringkasan Manusia Story 47 .....	44
Gambar 4. 12 Hasil Ringkasan Sistem Story 47 .....	45

## DAFTAR TABEL

Tabel 3. 1 Contoh <i>Case Folding</i> .....	17
Tabel 3. 2 Contoh <i>Remove Punctuation and Special Character</i> .....	18
Tabel 3. 3 Contoh <i>Stopword removal</i> .....	18
Tabel 3. 4 Contoh <i>Stemming</i> .....	19
Tabel 3. 5 menghitung jumlah kata dalam kalimat .....	20
Tabel 3. 6 Skor TF-ISF .....	24
Tabel 3. 7 Nilai Kemiripan Antar Kalimat.....	25
Tabel 3. 8 Hasil <i>Sentence Scoring</i> .....	25
Tabel 4. 1 Query Pembentukan Story multi dokumen.....	31
Tabel 4. 2 Contoh proses pelabelan .....	33
Tabel 4. 3. <i>Sentence Scoring</i> .....	34
Tabel 4. 4 Statistik Jumlah Kata Dalam Ringkasan <i>Fold 1</i> .....	36
Tabel 4. 5 Hasil Perhitungan ROUGE-1 <i>Fold 1</i> .....	37
Tabel 4. 6 Statistik Jumlah Kata Dalam Ringkasan <i>Fold 2</i> .....	38
Tabel 4. 7 Hasil Perhitungan ROUGE-1 <i>Fold 2</i> .....	39
Tabel 4. 8 Statistik Jumlah Kata Dalam Ringkasan <i>Fold 3</i> .....	40
Tabel 4. 9 Hasil Perhitungan ROUGE-1 <i>Fold 3</i> .....	40
Tabel 4. 10 Statistik Jumlah Kata Dalam Ringkasan <i>Fold 4</i> .....	42
Tabel 4. 11 Hasil Perhitungan ROUGE-1 <i>Fold 4</i> .....	42
Tabel 4. 12 Statistik Jumlah Kata Dalam Ringkasan <i>Fold 5</i> .....	43
Tabel 4. 13 Hasil Perhitungan ROUGE-1 <i>Fold 5</i> .....	44
Tabel 4. 14 Nilai Rata-Rata Nilai ROUGE-1 pada 5 <i>Fold</i> .....	46

## ABSTRAK

Fauzi, Deri. 2023. **Peringkasan Teks Multi Dokumen Berbahasa Indonesia Menggunakan *Sentence Scoring* dan *Support Vector Machine***. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Zainal Abidin, M.Kom. (II) Fatchurrochman, M.Kom.

Kata kunci: Peringkasan Teks, Multi Dokumen, SVM, Artikel *Online*.

Berita *online* berasal dari berbagai sumber portal berita yang tersedia secara luas di dunia maya. Namun, berita *online* yang melimpah dapat mengesampingkan detail dan keakuratan berita karena tujuannya untuk memberikan informasi terkini sebanyak mungkin. Ketersediaan berita *online* di internet dapat menyebabkan penerimaan informasi yang berlebihan, memberikan pemahaman yang kurang jelas mengenai substansi berita tersebut. Oleh karena itu, penting untuk menemukan representasi dokumen berita *online* guna memahami inti dari berita tersebut. Penelitian ini fokus pada menghasilkan ringkasan berita *online* multi dokumen dari ekstraksi fitur dan proses klasifikasi menggunakan *support vector machine*. penelitian ini mengklasifikan berita multi dokumen menggunakan ekstraksi fitur *Sentence Scoring* dan SVM. *Sentence Scoring* digunakan untuk input pada metode SVM agar dapat melakukan proses klasifikasi untuk menentukan hasil ringkasan. Hasil pengujian menunjukkan bahwa *Fold 3* memberikan hasil terbaik, dengan rata-rata *Recall* 0.946, *Precisi* 0.487, dan *F-Measure* 0.634. ROUGE-1 juga mencapai nilai tertinggi pada *Fold 3*, yaitu 0.946. Faktor kunci dalam hasil peringkasan adalah proses ekstraksi fitur menggunakan *Sentence Scoring* dan pelatihan data dengan SVM. Fitur seperti data numerik dan kemiripan antar kalimat berpengaruh signifikan terhadap hasil akhir dari peringkasan.

## ABSTRACT

Fauzi, Deri. 2023. **Multi Document Text Summarization in Indonesian Using *Sentence Scoring and Support Vector Machine***. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Dr. Zainal Abidin, M.Kom. (II) Fatchurrochman, M.Kom.

*Online* news originates from various news portal sources widely available on the internet. However, the abundance of *online* news may overlook details and accuracy in favor of providing the most up-to-date information. The availability of *online* news on the internet can lead to information overload, resulting in a vague understanding of the news substance. Therefore, it is crucial to find a representation of *online* news documents to comprehend the core of the news. This research focuses on generating multi-document *online* news summaries through feature extraction and classification processes using *support vector machines* (SVM). The study classifies multi-document news using *Sentence Scoring* feature extraction and SVM classification processes. *Sentence Scoring* is employed as input for the SVM method to perform the classification process to determine summary outcomes. Test results indicate that *Fold 3* yields the best results, with an average *Recall* of 0.946, *Precision* of 0.487, and *F-Measure* of 0.634. ROUGE-1 also achieves the highest value in *Fold 3*, which is 0.946. Key factors in the summarization results include the feature extraction process using *Sentence Scoring* and *training* data with SVM. Features such as numerical data and *Sentence* similarity significantly influence the final summarization outcomes

**Key words:** Text Summarization, Multi Document, SVM, *Online* News

## مستخلص البحث

فوزي ، دري. ٢٠٢٣ . تلخيص النصوص متعددة المستندات باللغة الإندونيسية باستخدام *Sentence Scoring and Support Vector Machine* رسالة ماجستير. قسم هندسة المعلوماتية ، كلية العلوم والتكنولوجيا ، جامعة مولانا ماليك إبراهيم الإسلامية الحكومية مالاغ. المشرفان: (١) الدكتور زين العابدين ، الماجستير (٢) فتح الرحمن ، الماجستير

الكلمات الأساسية: تلخيص النصوص ، مستندات متعددة ، آلة دعم المتجهات ، أخبار الإنترنت

تصدر الأخبار عبر الإنترنت من مصادر مختلفة من بوابات الأخبار المتوفرة على نطاق واسع في العالم الافتراضي. ومع ذلك ، يمكن أن تتجاهل الأخبار عبر الإنترنت الغنية التفاصيل والدقة للأخبار لأنها تهدف إلى تقديم أكبر قدر ممكن من المعلومات في أسرع وقت ممكن. يمكن أن تؤدي توفر الأخبار عبر الإنترنت على الإنترنت إلى قبول المعلومات المفرط ، مما يوفر فهمًا غير واضح للمضمون الإخباري. لذلك ، من المهم العثور على تمثيل لوثائق الأخبار عبر الإنترنت لفهم جوهر الأخبار. يركز هذا البحث على إنتاج ملخصات للأخبار عبر الإنترنت متعددة الوثائق من استخراج الميزات وعملية التصنيف باستخدام آلة دعم المتجهات. يقوم هذا البحث بتصنيف الأخبار متعددة الوثائق باستخدام استخراج الميزات *Sentence Scoring* و *SVM*. يتم استخدام *Sentence scoring* كمدخل لطريقة *SVM* حتى تتمكن من إجراء عملية التصنيف لتحديد نتيجة الملخص. أظهرت نتائج الاختبار أن الطية 3 قدمت أفضل النتائج ، بمتوسط استدعاء 0.946 ، ودقة 0.487 ، وقياس F 0.634. كما وصل ROUGE-1 إلى أعلى قيمة في الطية 3 ، وهي 0.946. العامل الرئيسي في نتيجة الملخص هو عملية استخراج الميزات باستخدام *Sentence scoring* وتدريب البيانات باستخدام *SVM*. تؤثر الميزات مثل البيانات الرقمية والتشابه بين الجمل بشكل كبير على النتيجة النهائية للملخص.



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan internet pada saat ini berlangsung dengan cepat yang memungkinkan akses mudah bagi setiap individu di berbagai waktu dan lokasi. Kemudahan ini memungkinkan seseorang untuk dengan cepat membuat dan mengakses informasi dalam jumlah yang signifikan. Contoh nyata dari fenomena ini adalah melalui ketersediaan berita *online* yang melimpah di platform internet. Berita *online* tersebut berasal dari berbagai sumber portal berita yang tersedia secara luas di dunia maya. Namun, berita *online* yang melimpah tersebut dapat mengesampingkan detail dan keakuratan berita karena tujuannya untuk memberikan informasi terkini sebanyak mungkin. Ketersediaan berita *online* di internet dapat menyebabkan penerimaan informasi yang berlebihan, memberikan pemahaman yang kurang jelas mengenai substansi berita tersebut. Pada beberapa kesempatan, seseorang cenderung membaca berita dari satu situs ke situs lainnya untuk membandingkan konten yang berkaitan dengan topik tertentu. Proses ini tidak hanya memakan waktu secara signifikan, tetapi juga sering kali menghadirkan kalimat-kalimat yang memiliki makna serupa, mengakibatkan pembaca melakukan tugas yang serupa berulang-ulang (Hayatin *et al.*, 2018). Oleh karena itu, penting untuk menemukan representasi dari dokumen berita *online* untuk memahami tentang isi berita *online* tersebut.

Representasi dari dokumen berita *online* bisa disebut juga dengan ringkasan. Ringkasan merupakan hal yang penting untuk memahami suatu dokumen atau teks.

Dengan membuat ringkasan, maka seseorang dapat lebih mudah untuk memahami informasi yang terkandung didalamnya. Agar dapat menghasilkan ringkasan yang efektif dan mudah dipahami, seseorang perlu memiliki pemahaman yang memadai terhadap topik yang terkandung dalam teks atau dokumen tersebut. Apabila telah mencapai tingkat pemahaman yang memadai terhadap teks atau dokumen, maka penyusunan ringkasan biasanya menjadi lebih mudah dilakukan. Walau demikian, perlu dicatat bahwa ringkasan yang dihasilkan oleh setiap individu kemungkinan besar akan menunjukkan perbedaan, mengingat hal tersebut merupakan hasil dari pemikiran subjektif masing-masing individu. Dalam konteks penciptaan ringkasan yang efektif dan dapat dipahami secara menyeluruh, penting untuk melakukan pertimbangan secara objektif dan cermat. Untuk membuat ringkasan dari sebuah dokumen teks, dibutuhkan pemahaman dan pemikiran dari pembaca terhadap maksud penulis dokumen tersebut. Dengan memahami kedua aspek tersebut, ada cara untuk mengekstrak inti dari dokumen tersebut, yaitu dengan cara meringkasnya. Proses meringkas teks dokumen ini cukup melelahkan bagi manusia, terlebih jika dokumen yang diringkas sangat panjang. Oleh karena itu diperlukan suatu sistem yang mampu melakukan peringkasan teks atau *Text Summarization*.

*Text Summarization* atau peringkasan teks secara otomatis adalah bagian dari bidang *Natural Language Processing (NLP)* yang tujuan utamanya adalah membuat ringkasan berdasarkan kalimat-kalimat yang relevan dalam dokumen asli (Widodo *et al.*, 2021). Peringkasan teks otomatis dapat membantu manusia untuk mendapatkan informasi utama dari dokumen asli dengan cepat tanpa perlu membaca isi keseluruhan dokumen. Dalam konteks ini, peringkasan menjadi alat

yang strategis untuk menyajikan inti berita tanpa memerlukan waktu yang berlebihan, memudahkan pembaca dalam memahami peristiwa terkini tanpa harus membaca seluruh artikel. Dengan peringkasan, pembaca dapat dengan mudah memahami esensi informasi tanpa harus terjebak dalam detail yang berlebihan, mengoptimalkan pengalaman membaca *online* mereka. Terlebih lagi, dalam era di mana waktu menjadi salah satu hal yang berharga, peringkasan menjadi alat yang tidak hanya memberikan kemudahan tetapi juga meningkatkan efisiensi dalam pemahaman berita, mendukung pembaca dalam membuat keputusan yang informasional. Dengan demikian, peringkasan artikel berita *online* bukan hanya respons terhadap kebutuhan pembaca, tetapi juga menjawab tuntutan zaman yang menekankan kecepatan dan efisiensi dalam akses informasi.

Dalam agama Islam, perlu sikap *tabayyun* untuk pengecekan ulang tentang suatu informasi agar terhindar dari berita yang tidak sesuai dengan judulnya. Dalam peringkasan ini kita tidak akan tertipu dengan judul berita karena isi dari ringkasan merupakan hasil analisis dari suatu berita, tidak hanya berdasarkan judulnya saja. Allah S.W.T berfirman pada surat Al-Hujurat ayat 6:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِنْ جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَنْ تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْحَبُوا عَلَىٰ مَا فَعَلْتُمْ نَادِمِينَ

“Hai orang-orang yang beriman, jika datang kepadamu orang fasik membawa suatu berita, maka periksalah dengan teliti agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu.” (Q.S Al-Hujurat: 6 )

Tafsir jalalain memberikan penjelasan mengenai ayat ini yaitu (Hai orang-orang yang beriman! Jika datang kepada kalian orang fasik membawa suatu berita) (maka periksalah oleh kalian) kebenaran beritanya itu, apakah ia benar atau

berdusta. Menurut suatu qiraat dibaca Fatatsabbatuu berasal dari lafal Ats-Tsabaat, artinya telitilah terlebih dahulu kebenarannya (agar kalian tidak menimpakan musibah kepada suatu kaum) menjadi Maf'ul dari lafal Fatabayyanuu, yakni dikhawatirkan hal tersebut akan menimpa musibah kepada suatu kaum (tanpa mengetahui keadaannya) menjadi Hal atau kata keterangan keadaan dari Fa'il, yakni tanpa sepengetahuannya (yang menyebabkan kalian) membuat kalian (atas perbuatan kalian itu) yakni berbuat kekeliruan terhadap kaum tersebut (Suyutī & Mahalli, 2010).

Oleh karena itu, penting untuk menyaring berita sehingga kita tidak jatuh pada ketidaktahuan. Demikian pula, artikel berita yang panjang telah diringkas sehingga pembaca dapat memperoleh banyak informasi tentang beberapa topik dalam waktu yang singkat. Ada beberapa pendekatan untuk meringkas teks secara otomatis dan salah satunya adalah dengan menggunakan pendekatan *machine learning*. Dalam penelitian yang dilakukan oleh (Rahul *et al.*, 2020), dalam lima tahun terakhir metode yang paling umum ditemukan untuk peringkasan teks otomatis adalah menggunakan pendekatan *supervised learning*. Pendekatan ini menggunakan algoritma klasifikasi seperti *Naïve Bayes*, *Decision Tree*, *K-nearest Neighbor*, dan *Support Vector Machine* (SVM) dan lain sebagainya.

Pada penelitian ini menggunakan metode *Support Vector Machine* (SVM) dengan *Sentence Scoring* pada setiap kalimatnya. *Sentence Scoring* perlu dilakukan karena untuk representasi awal agar mesin dapat membedakan setiap kalimat yang terdapat di dalamnya. Setiap fitur dalam penilaian kalimat akan dijadikan input awal untuk algoritma SVM, yang akan mencari pola atau pengetahuan dari data

tersebut. Dengan menggunakan metode ini, proses peringkasan teks akan dianggap sebagai sebuah masalah klasifikasi kalimat. Setiap kalimat akan diberi bobot atau nilai dengan *Sentence Scoring* untuk mendapatkan nilai input dari SVM. Kemudian akan dilakukan klasifikasi kalimat untuk mendapatkan hasil ringkasan.

## **1.2 Pernyataan Masalah**

Dengan latar belakang yang telah dipaparkan, maka pernyataan masalah pada penelitian ini yaitu, berapa nilai relevansi dari hasil peringkasan teks multi dokumen berbahasa Indonesia menggunakan algoritma *Support Vector Machine* (SVM) jika diukur menggunakan ROUGE

## **1.3 Tujuan Penelitian**

Tujuan yang ingin dicapai melalui penelitian ini adalah untuk mengukur tingkat relevansi *ROUGE* dari peringkasan teks multi dokumen berbahasa Indonesia menggunakan algoritma *Support Vector Machine* .

## **1.4 Batasan Masalah**

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian yang akan dilakukan adalah mengimplementasikan algoritma SVM untuk peringkasan teks multi dokumen berbahasa Indonesia dengan pendekatan ekstraktif
2. Jenis dokumen yang akan diringkas dalam penelitian ini adalah dokumen berita *online*
3. Dataset yang digunakan dalam penelitian ini diambil dari kumpulan berita *online* berbahasa Indonesia INDOSUM.

## **1.5 Manfaat Penelitian**

Penelitian ini akan membuktikan tingkat akurasi dan hasil peringkasan multi dokumen dengan algoritma SVM. Dengan mengetahui hal tersebut, metode SVM dapat dibandingkan dengan metode peringkasan teks lainnya. Sistem peringkasan otomatis ini juga bermanfaat bagi jurnalis dan pembaca berita *online* karena dapat menyediakan informasi yang mudah dicari dan memudahkan dalam mendapatkan informasi secara cepat dan efisien.

## **BAB II**

### **STUDI PUSTAKA**

#### **2.1 *Text Summarization***

*Text Summarization* atau peringkasan teks otomatis adalah proses membuat teks yang Panjang menjadi berkurang dengan bantuan program dari komputer sehingga dihasilkan sebuah ringkasan yang berisikan kalimat-kalimat penting sehingga dengan menggunakan peringkasan teks otomatis akan mendapatkan ide pokok maupun informasi penting dari sebuah artikel *online* secara jelas, tepat, dan ringkas, tanpa mengubah makna dari isi artikel tersebut. Menurut Allahyari (2017) Peringkasan teks otomatis adalah tugas menghasilkan ringkasan dan ringkasan yang lancar sambil mempertahankan konten informasi utama dan makna keseluruhan.

Penelitian tentang *Text Summarization* atau peringkasan teks pertama kali dilakukan oleh Luhn (1958) yang melakukan peringkasan secara otomatis pada abstrak artikel jurnal, artikel majalah dan makalah teknis menggunakan metode statistik. Penelitian Luhn ini memiliki dampak yang penting untuk penelitian tentang peringkasan teks otomatis kedepannya. Setelah penelitian yang dilakukan oleh Luhn, muncul jenis teknik pada peringkasan teks otomatis, yaitu *Extractive Summarization*.

*Extractive Summarization* atau peringkasan secara ekstraktif adalah teknik untuk menghasilkan ringkasan dengan memilih bagian dari himpunan kalimat dalam teks aslinya. Menurut Moratanch & Chitrakala (2017) Dalam teknik peringkasan ekstraktif, dilakukan seleksi kalimat atau paragraf penting dari naskah

asli dan disusun menjadi sebuah bentuk yang lebih singkat. Pentingnya sebuah kalimat ditentukan oleh ciri-ciri statistik dan linguistik dari kalimat tersebut.

Pada penelitian yang dilakukan oleh Christian (2016) bertujuan untuk mengembangkan metode peringkasan otomatis teks berdasarkan TF-IDF. Dalam penelitian ini, Christian memakai metode ekstraktif untuk menemukan kalimat-kalimat penting dalam satu dokumen dan memilihnya sebagai bagian dari ringkasan. Hasil penelitian menunjukkan hasil tertinggi dari *F-Measure* sebesar 0,666.

Penelitian tentang teknik ekstraktif juga dilakukan oleh Widodo (2021) untuk dokumen berbahasa Indonesia. Dalam penelitian ini, dilakukan perbandingan performa beberapa algoritma yang umum digunakan dalam peringkasan teks ekstraktif, seperti *TextRank*, *LexRank*, dan *LSA*, pada bahasa Indonesia. Ditemukan bahwa algoritma *Sentence Scoring dan decision tree* memiliki kinerja yang lebih baik dalam melakukan peringkasan teks pada bahasa Indonesia jika dibandingkan dengan *TextRank*, *LexRank* dan *LSA* dengan nilai *F-score* sebesar 0.58.

Pendekatan *machine learning* juga digunakan dalam penelitian yang dilakukan oleh Belwal (2022) tentang peringkasan dokumen ekstraktif dengan topic modelling berbasis klaster yang mengelompokkan kalimat berdasarkan topik yang sama untuk menentukan kalimat mana yang penting dan relevan untuk diambil sebagai bagian dari ringkasan. Metode yang diusulkan oleh Belwal berhasil menghasilkan ringkasan yang lebih baik dibandingkan dengan metode-metode ekstraktif lainnya dengan nilai *Precision* sebesar 0,156 dan nilai *Recall* sebesar 0.453



Untuk pendekatan *machine learning* dengan Bahasa Indonesia juga dilakukan oleh Aulia (2021) dalam penelitiannya, Aulia bertujuan mengembangkan algoritma ringkasan ekstraktif yang efektif untuk bahasa Indonesia dengan menggunakan *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa algoritma SVM yang diusulkan mampu menghasilkan ringkasan yang lebih baik dibandingkan metode ekstraktif lainnya dengan hasil akurasi tertinggi diperoleh pada skenario 1, dengan data *training* 70% dan data *testing* 30%, untuk linear kernel yaitu Accuracy 72%, *Presicion* 63%, *Recall* 51% dan *F-Measure* 57%.

## **2.2 Text Preprocessing**

*Preprocessing* merupakan proses yang penting dalam melakukan klasifikasi teks. Tahapan *Preprocessing* dapat mempengaruhi hasil dari proses klasifikasi. Dalam melakukan proses *Preprocessing* ini terdapat beberapa tahapan yaitu Proses *Preprocessing* meliputi *unitization* dan *tokenization*, standardisasi dan *cleansing* atau pembersihan data teks, *stopword removal*, dan *stemming* atau lemmatisasi (Anandarajan *et al.*, 2018).

Menurut Naufal Ammar & Suyanto (2020) untuk melakukan peringkasan teks bahasa Indonesia, diperlukan tahapan *praprocessing* yang mencakup segmentasi (pemisahan teks dokumen menjadi sekumpulan kalimat), *tokenisasi* (pemisahan kalimat menjadi sekumpulan kata), menghapus *stopword* atau kata yang tidak penting, *stemming* (pengelompokan kumpulan kata menjadi kata dasar), dan terakhir pembuatan dokumen yang berisi daftar kata yang telah diproses dengan *stemming*.

### 2.3 *Sentence Scoring*

*Sentence Scoring* atau penilaian kalimat merupakan proses untuk memberikan nilai pada setiap kalimat. Dalam pendekatan peringkasan teks ekstraktif, setiap kalimat dalam teks diberi bobot yang unik melalui proses *Sentence Scoring*. Pendekatan ini menggunakan metode berbasis statistik, sehingga tidak memerlukan pengetahuan linguistik khusus dan dapat diterapkan pada berbagai bahasa (*language independent*). Beberapa fitur statistik yang digunakan meliputi posisi kalimat, panjang kalimat, dan keberadaan data numerik, (Gambhir & Gupta, 2017).

Pada penelitian Qaroush *et al.* (2021) menggunakan Metode ekstraktif yang diusulkan untuk mengevaluasi setiap kalimat berdasarkan kombinasi fitur statistik dan semantik di mana formulasi baru digunakan dengan mempertimbangkan kepentingan, cakupan, dan keragaman kalimat. Kemudian Qaroush menggunakan algoritma klasifikasi untuk mengetahui hasil dari perumusan fitur diatas menggunakan algoritma klasifikasi seperti Naive Bayes, SVM (dengan kernel RBF), jaringan saraf dua lapis, J48, dan Random Forest. Hasil penelitian menunjukkan kelebihan metode yang diusulkan dalam hal metrik kinerja presisi, *Recall*, dan *F-score*.

Metode *machine learning* menganggap peringkasan teks sebagai permasalahan pada klasifikasi yaitu dengan memprediksi apakah sebuah kalimat “layak” untuk dijadikan sebagai ringkasan akhir atau tidak. Seperti pada penelitian yang dilakukan oleh Somantri *et al.* (2018) yaitu dengan melakukan peringkasan teks otomatis berdasarkan klasifikasi kalimat menggunakan SVM. Proses *Sentence*

*Scoring* pada penelitian ini meliputi ekstraksi fitur pada panjang kalimat, posisi kalimat, jumlah data numerik, TF-ISF, dan kemiripan antar kalimat. Hasil dari ekstraksi fitur tersebut akan menjadi inputan untuk membuat model SVM .

#### **2.4 Support Vector Machine**

*Support vector machine* pertama kali diperkenalkan oleh Vapnik (1999) tentang pembelajaran statistik dan aplikasinya pada pembelajaran mesin (*machine learning*). Penelitian Vapnik ini menjadi salah satu penelitian yang paling banyak dikutip dalam bidang pembelajaran mesin dan menjadi landasan bagi banyak algoritma pembelajaran mesin modern seperti *Support Vector Machine* (SVM). Metode ini adalah salah satu metode pada *machine learning* yang termasuk pada *supervised learning* yang menganalisis dan mengenali suatu pola tertentu.

*Support vector machine* dapat digunakan untuk pengolahan text processing seperti analisis sentimen, klasifikasi berita, dan peringkasan teks. Pada penelitian Saragih *et al.* (2021) tentang analisis sentimen tentang PPKM di Jakarta pada media sosial Twitter. Dari penerapan metode *Support vector machine* terhadap penelitian tersebut dihasilkan tingkat akurasi sebesar 85%.

Penelitian menggunakan metode klasifikasi juga dilakukan oleh Yovellia Londo *et al.* (2019) pada penelitiannya tentang klasifikasi teks berita menggunakan metode *supervised learning*. Penelitian ini membandingkan algoritma klasifikasi pada *machine learning* untuk mengklasifikasikan berita berbahasa Indonesia. Dari hasil penelitian tersebut dihasilkan metode *support vector machine* memperoleh tingkat akurasi paling besar di antara metode klasifikasi lainnya, yaitu sebesar 93%.

Somantri *et al.* (2018) melakukan penelitian tentang peringkasan teks berbahasa indonesia menggunakan *support vector machine*. Hasil penelitian menunjukkan bahwa *support vector machine* dapat digunakan untuk peringkasan teks otomatis dengan nilai *Recall* 42% , *Presicion* 60% , dan *F-Measure* 50%

Penelitian lain dilakukan oleh Qaroush *et al.* (2021) tentang peringkasan teks otomatis berbahasa arab dengan kombinasi fitur statistik dan semantik. Penelitian ini menggunakan ekstraksi fitur berbasis skor dan model klasifikasi pada *machine learning*. Hasil penelitian ini menunjukkan bahwa kombinasi *score-based* dan algoritma *support vector machine* mendapatkan nilai *F-score* terbesar dengan nilai 0,554 pada ROUGE-1 dan ROUGE-2 dibandingkan dengan algoritma klasifikasi lainnya.

SVM diciptakan untuk memberikan pendekatan yang fleksibel dalam menentukan batas keputusan antara dua kelas. Margin adalah area terbesar yang digunakan untuk memisahkan kedua kelas. Dalam menemukan *support vector*, titik data atau *dot product* yang paling dekat dengan margin sangat penting. Agar margin dapat diperluas dengan efektif, teknik *kernel trick* digunakan dengan menambahkan dimensi ekstra ke margin melalui matriks kernel. Salah satu kernel yang ada di SVM adalah kernel linear. Kernel linear dapat diformulasikan seperti berikut ini,

$$K(x_i, x_j) = x_i x_j \quad (2,1)$$

Kernel linear merupakan fungsi kernel yang dapat menghitung hasil perkalian titik antara dua vektor pada ruang fitur yang sama tanpa perlu melakukan perubahan dimensi. Dalam SVM, kernel linear berfungsi sebagai alat untuk memisahkan dua kelas data dengan menggunakan *hyperplane* (bidang pemisah) yang memiliki

margin maksimum pada ruang fitur. Kernel linear menjadi salah satu jenis kernel yang paling sering digunakan dalam SVM karena kemampuannya yang cukup efektif dalam memproses data dengan dimensi yang relatif besar. Dengan menggunakan kernel linear, SVM dapat memproses data dengan lebih efisien dan menghasilkan model yang lebih sederhana serta mudah diinterpretasikan.

## 2.5 ROUGE

ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) adalah metode evaluasi untuk mengukur performa hasil dari peringkasan teks. ROUGE adalah suatu teknik pengukuran secara otomatis untuk mengevaluasi kualitas dari sebuah ringkasan. Teknik ini bekerja dengan cara membandingkan kemiripan antara ringkasan yang dihasilkan oleh mesin dengan ringkasan yang dibuat oleh manusia sebagai acuan yang benar. Dalam hal ini, semakin mirip ringkasan mesin dengan ringkasan manusia, maka semakin tinggi pula nilai ROUGE yang diperoleh dan kualitas ringkasan yang dihasilkan dianggap semakin baik (Verma *et al.*, 2019). Persamaan untuk mengevaluasi hasil ringkasan dengan ROUGE adalah sebagai berikut

$$\text{ROUGE} - N = \frac{\sum_{S \in \text{Sum}_{ref}} \sum_{N\text{-gram} \in S} \text{Count}_{\text{match}}(N - \text{gram})}{\sum_{S \in \text{Sum}_{ref}} \sum_{N\text{-gram} \in S} \text{Count}(N - \text{gram})} \quad (2.2)$$

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Pengumpulan Data

Data diambil dari kumpulan artikel *online* berbahasa Indonesia Bernama INDOSUM. Dataset ini berisi artikel berbahasa indonesia dari berbagai sumber berita *online* beserta ringkasannya (Kurniawan & Louvan, 2018). Dataset yang digunakan hanya beritanya saja dikarenakan pada dataset Indosum terdiri dari data untuk single dokumen. Pada penelitian ini menggunakan multi dokumen yang dibentuk dari berita single dokumen yang ada pada dataset Indosum tersebut.

category	gold_labels	id	paragraphs	source	source_url	summary
0	tajuk utama [[False, True], [True, True], [False, False, F...]]	1501893029-lula-kamal-dokter-nyan-thamrin-saki...	[[[Jakarta, , CNN, Indonesia, -, -, Dokter, R...]]	cnn indonesia	https://www.cnnindonesia.com/hiburan/201708041...	[[Dokter, Lula, Kamal, yang, merupakan, selebr...]]
1	teknologi [[False, False, False, False], [False, True, T...]]	1509072914-dua-smartphone-zenfone-baru-tawarka...	[[[Selfie, ialah, salah, satu, tema, terpanas, ...]]	dailysocial.id	https://dailysocial.id/post/dua-smartphone-zen...	[[Asus, memperkenalkan, , ZenFone, generasi, ...]]
2	hiburan [[True], [True], [False, False], [False], [Fal...]]	1510613677-songsong-visti-2020-bengkulu-perkua...	[[[Jakarta, , CNN, Indonesia, -, -, Dinas, Pa...]]	cnn indonesia	https://www.cnnindonesia.com/gaya-hidup/201711...	[[Dinas, Pariwisata, Provinsi, Bengkulu, kempa...]]
3	tajuk utama [[True, True], [False, False, False], [True], ...]]	1502706803-icw-ada-kejanggalaan-atas-lewasnya-s...	[[[Merdeka.com, -, Indonesia, Corruption, Watc...]]	merdeka	https://www.merdeka.com/peristiwa/icw-merasa-a...	[[Indonesia, Corruption, Watch, (, ICW, ), mem...]]
4	tajuk utama [[False, True], [True, True, True], [False], [...]]	1503039338-pembagian-sepeda-usal-upacara-penur...	[[[Merdeka.com, -, Presiden, Joko, Widodo, (, ...]]	merdeka	https://www.merdeka.com/peristiwa/usal-upacara...	[[Jokowi, memimpin, upacara, penurunan, bender...]]
...	...	...	...	...	...	...
14257	tajuk utama [[True], [True], [False], [False, False], [Fal...]]	1497645345-as-kirimkan-peluncuran-ke-suria...	[[[Jakarta, , CNN, Indonesia, -, -, Amerika, ...]]	cnn indonesia	http://www.cnnindonesia.com/internasional/2017...	[[Amerika, Serikat, dilaporkan, telah, mengiri...]]
14258	olahraga [[True, True], [True, False], [False], [False, ...]]	1495406700-borneo-bersyukur-tahan-persib-di-gola...	[[[Bandung, , CNN, Indonesia, -, -, Borneo, F...]]	cnn indonesia	http://www.cnnindonesia.com/olahraga/201705202...	[[Borneo, FC, menahan, Imbang, Persib, Bandung...]]
14259	tajuk utama [[True], [False, False], [True, True, False], ...]]	1513941815-mantan-dijen-perhubungan-laut-sege...	[[[JAKARTA, (, Pos, Kota, ), -, Komisi, Pember...]]	poskotanews	http://poskotanews.com/2017/12/21/mantan-dirje...	[[Komisi, Pemberantasan, Korupsi, (, KPK, ), s...]]
14260	olahraga [[True, True], [True, True], [False, False], [...]]	1496440800-rangking-ffa-indonesia-naik-dua-pe...	[[[Merdeka.com, -, Sebuah, kabar,embira, dat...]]	merdeka	https://www.merdeka.com/sepakbola/rangking-ff...	[[Kabar,embira, datang, bagi, , sepakbola, ...]]
14261	teknologi [[True, False], [False, False, False, False, F...]]	1505421900-apple-tv-4k-usung-resolusi-super-be...	[[[Ada, satu, perangkat, menarik, lainnya, yan...]]	dailysocial.id	https://dailysocial.id/post/apple-tv-4k-hdr	[[Ada, satu, perangkat, menarik, lainnya, yang...]]

14262 rows x 7 columns

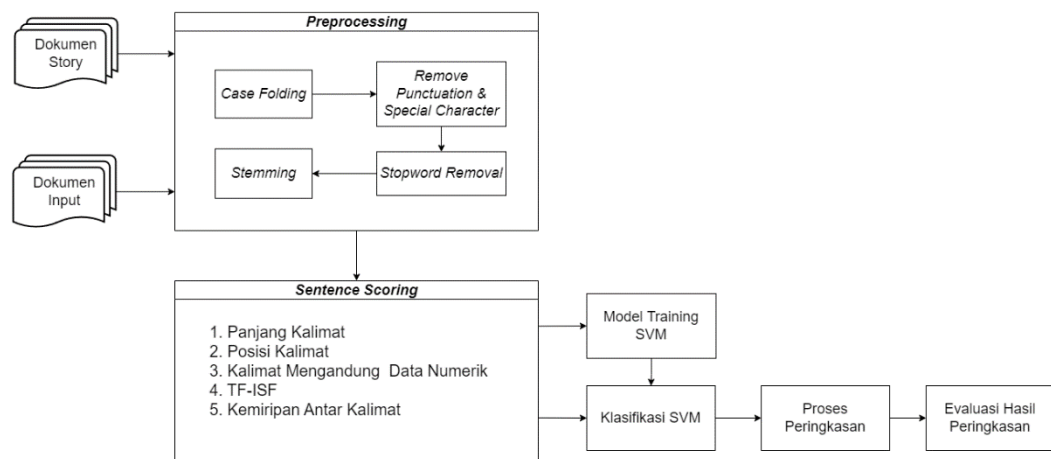
Gambar 3. 1 Data dari INDOSUM

Seperti pada Gambar 3. 1 diatas, jumlah total dataset INDOSUM sebanyak 14 ribu lebih berita *online* yang diperoleh dari berbagai sumber. Kumpulan berita dari dataset ini dibagi menjadi beberapa kategori seperti tajuk utama, teknologi, olahraga, hiburan, dan showbiz. Kemudian waktu terbit berita yang ada di dataset ini berkisar antara tahun 2016-2018. Pada penelitian ini tidak menggunakan

keseluruhan atribut dataset tersebut, akan tetapi hanya mengambil berita dan kategori saja.

### 3.2 Desain Sistem

Desain sistem pada Gambar 3. 2 meliputi pembuatan story multi dokumen, *Preprocessing*, *Sentence Scoring*, normalisasi, pembuatan model klasifikasi, dan evaluasi hasil dari model SVM.



Gambar 3. 2 Desain Sistem

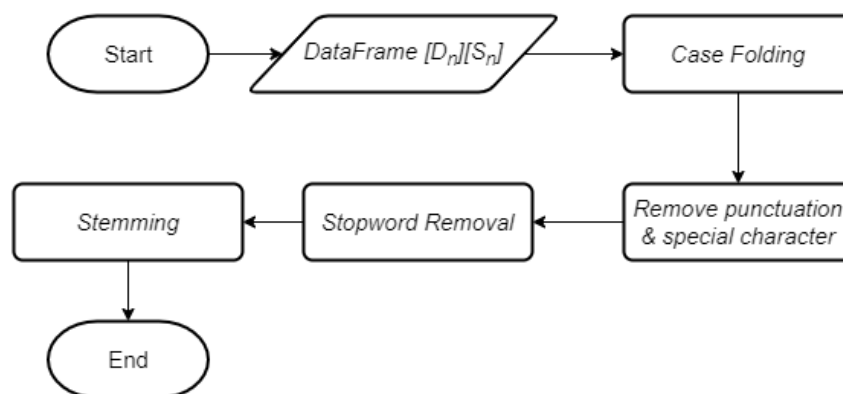
Masing-masing proses diatas dapat diuraikan sebagai berikut: 1) Secara garis besar proses peringkasan multi dokumen dengan *support vector machine* terbagi dalam dua bagian yaitu proses *training* data dan *testing* data. 2) Masing-masing kalimat pada dataset akan dilakukan praprocessing dengan melakukan proses yang ada pada Gambar 3. 3) selanjutnya, masing-masing kalimat akan diekstrak fiturnya dengan lima fitur yang ada didalam *Sentence Scoring* sehingga masing-masing kalimat akan mendapatkan nilai fitur yang membedakan kalimat tersebut dengan kalimat lainnya. 4) *Support vector machine* akan melakukan proses

klasifikasi pada setiap kalimat secara bersamaan dengan memperhitungkan fitur-fitur yang dimilikinya untuk menghasilkan output dalam bentuk *predicted* label. 5) untuk data *testing* dengan dokumen input, alur proses yang dilakukan sama, mulai dari *preprocessing* sampai *Sentence Scoring*, akan tetapi setelah masuk ke SVM, data *testing* akan diklasifikasi menggunakan model SVM yang sudah dibuat sebelumnya. Setelah itu dilakukan proses peringkasan dengan mengambil label-label serta kalimat yang akan dimasukkan kedalam ringkasan.

### 3.3 Preprocessing

Pada tahap ini kalimat akan di preprocessing meliputi *case folding* untuk mengubah huruf dalam kalimat menjadi huruf kecil, *remove punctuation and special character* untuk menghapus tanda baca dan karakter special pada kalimat, *stopword removal* untuk menghapus kata-kata umum yang sering muncul pada kalimat, dan *stemming* untuk mengubah kata dalam kalimat menjadi kata dasar. Proses ini dilakukan untuk mendapatkan data yang bersih dan dapat digunakan sebagai input pada tahap *Sentence Scoring* dan SVM (Anandarajan *et al.*, 2018).

Tahapan pada *Preprocessing* dapat dilihat pada Gambar 3. 3



Gambar 3. 3 Flowchart Praprocessing



Pada umumnya, praprocessing pada NLP melibatkan tahapan tokenizing atau memisahkan kalimat menjadi token / kata, namun pada dataset yang digunakan yaitu dataset INDOSUM, setiap kalimat sudah ditokenisasi sehingga tidak perlu lagi melakukan tahapan tokenisasi.

### 3.3.1 Case Folding

*Case Folding* yaitu tahapan awal dalam text *Preprocessing* yaitu mengubah semua huruf dalam kalimat menjadi huruf kecil untuk menghindari perbedaan yang tidak perlu dalam teks.

Tabel 3. 1 Contoh *Case Folding*

Sebelum <i>Case Folding</i>	["Himpunan", "Mahasiswa", "Islam", "(", "HMI", ")", "mendesak", "pemerintah", "memberikan", "perlindungan", "hukum", "dan", "keamanan", "kepada", "tokoh", "-", "tokoh", "di", "daerah", "."]
Setelah <i>Case Folding</i>	["himpunan", "mahasiswa", "islam", "(", "hmi", ")", "mendesak", "pemerintah", "memberikan", "perlindungan", "hukum", "dan", "keamanan", "kepada", "tokoh", "-", "tokoh", "di", "daerah", "."]

### 3.3.2 Remove Punctuation and Special Character

Remove punctuation adalah tahapan selanjutnya dari *preprocessing*. Proses ini menghapus semua tanda baca pada kalimat. Tanda baca yang dimaksud seperti titik, koma, tanda tanya, dan tanda seru dihapus karena tidak bermakna dan tidak penting untuk analisa teks dan dapat mengganggu hasil analisis. Kemudian special character seperti tanda kurung, simbol matematika, dan karakter khusus lainnya juga dihapus karena tidak memiliki peran penting dalam analisis teks dan dapat mengganggu hasil analisis.

Tabel 3. 2 Contoh *Remove Punctuation and Special Character*

Sebelum <i>Remove Punctuation</i>	["himpunan", "mahasiswa", "islam", "(", "hmi", ")", "mendesak", "pemerintah", "memberikan", "perlindungan", "hukum", "dan", "keamanan", "kepada", "tokoh", "-", "tokoh", "di", "daerah", "."]
Setelah <i>Remove Punctuation</i>	["himpunan", "mahasiswa", "islam", "hmi", "mendesak", "pemerintah", "memberikan", "perlindungan", "hukum", "dan", "keamanan", "kepada", "tokoh", "tokoh", "di", "daerah"]

### 3.3.3 *Stopword Removal*

*Stopword removal* merupakan sebuah proses dimana kata-kata umum atau yang disebut dengan *stopword* akan dihilangkan dari sebuah teks karena kata-kata tersebut tidak memberikan kontribusi penting dalam analisis teks. Contohnya, kata-kata seperti kata depan, kata sambung, dan kata-kata umum lainnya. Proses *stopword removal* dilakukan agar proses analisis teks dapat berjalan lebih cepat dan juga dapat mengurangi dimensi data yang tidak perlu, sehingga dapat meningkatkan kualitas dari analisis yang dilakukan. Contoh dari kata yang dihapus adalah "di", "ke", "yang", "dan", dan sebagainya.

Tabel 3. 3 Contoh *Stopword removal*

Sebelum <i>Stopword removal</i>	["himpunan", "mahasiswa", "islam", "hmi", "mendesak", "pemerintah", "memberikan", "perlindungan", "hukum", "dan", "keamanan", "kepada", "tokoh", "tokoh", "di", "daerah"]
Setelah <i>Stopword removal</i>	["himpunan", "mahasiswa", "islam", "hmi", "mendesak", "pemerintah", "memberikan", "perlindungan", "hukum", "keamanan", "tokoh", "tokoh", "daerah"]

### 3.3.4 *Stemming*

Proses terakhir pada text processing adalah *stemming*, yang mana merujuk pada proses transformasi kata dalam teks menjadi bentuk dasar atau kata dasar yang memiliki makna yang sama. Dengan melalui proses ini, kata-kata dalam teks akan

memiliki representasi yang lebih konsisten dan mudah diproses pada tahapan berikutnya.

Tabel 3. 4 Contoh *Stemming*

Sebelum <i>Stemming</i>	["himpunan", "mahasiswa", "islam", "hmi", "mendesak", "pemerintah", "memberikan", "perlindungan", "hukum", "keamanan", "tokoh", "tokoh", "daerah"]
Setelah <i>Stemming</i>	["himpun", "mahasiswa", "islam", "hmi", "desak", "perintah", "beri", "lindung", "hukum", "aman", "tokoh", "tokoh", "daerah"]

### 3.4 Implementasi *Sentence Scoring*

Tahap *sentence scoring* adalah memberikan skor pada setiap kalimat agar memiliki nilai. Terdapat 5 fitur yang digunakan pada tahap ini yaitu Panjang kalimat, Posisi kalimat, kalimat mengandung data numerik, *TF-ISF* (*Term Frequency-Inverse Sentence Frequency*), dan kemiripan kalimat yang dihitung dengan *cosine similarity*. Tujuan dilakukan *sentence scoring* adalah untuk proses awal menentukan keunikan dari setiap kalimatnya (Sabuna & Setyohadi, 2018). Setiap fitur dalam *sentence scoring* akan menjadi input awal untuk algoritma SVM. Algoritma ini akan menemukan pola dari data tersebut. Pola ini kemudian digunakan oleh sistem untuk menentukan kalimat mana yang akan dimasukkan ke dalam ringkasan.

#### 3.4.1 Panjang Kalimat (F1)

Dalam memilih proses ringkasan kalimat, panjang kalimat menjadi salah satu pertimbangan yang penting. Kandidat kalimat ringkasan dipilih berdasarkan panjang kalimat, dimana kalimat terpanjang dijadikan sebagai referensi (Qaroush *et al.*, 2021). Berikut contoh dari fitur Panjang kalimat pada sebuah dokumen

*"Himpunan Mahasiswa Islam (HMI) mendesak pemerintah memberikan perlindungan hukum dan keamanan kepada tokoh-tokoh di daerah. Pasalnya, selama ini, tokoh-tokoh di daerah kerap menjadi sasaran kekerasan. Tuntutan itu disampaikan Ketua Umum Pengurus Besar HMI M. Fachruddin, saat bertemu pimpinan Fraksi Utusan Daerah MPR, di Gedung DPR/MPR, Senayan, Jakarta Selatan, Selasa (15/5). Menurut Fachruddin, ketegangan politik yang terjadi di pusat kerap mengorbankan tokoh-tokoh di daerah, terutama di daerah yang sedang bergolak. Sejumlah tokoh yang menjadi korban kekerasan adalah Rektor Institut Agama Islam Negeri Arraniry Aceh Syofwan Idris, Tengku Nashiruddin Daud, anggota DPR asal Aceh dan Jafar Sidik Hamzah, Ketua International Forum for Aceh.*

Gambar 3. 4 contoh berita yang akan di *preprocessing*

Untuk menghitung panjang kalimat, jumlah kata dalam sebuah kalimat dibagi dengan jumlah kata dalam kalimat terpanjang. Cara menghitung panjang kalimat dapat dilihat pada persamaan berikut ini.

$$F1 = \frac{\# \text{ kalimat } j}{\# \text{ kata dalam kalimat terpanjang}} \quad (3.1)$$

Dokumen diatas akan dibagi perkalimat dan dihitung jumlah kata pada masing-masing kalimat. Hasilnya dapat dilihat pada Tabel 3. 5 dibawah ini

Tabel 3. 5 menghitung jumlah kata dalam kalimat

$S_n$	Kalimat setelah di <i>praprocessing</i>	n-kata
1	<i>himpun mahasiswa islam hmi desak perintah beri lindung hukum aman tokoh tokoh daerah</i>	13
2	<i>pasal lama tokoh tokoh daerah kerap jadi sasar keras</i>	9
3	<i>tuntut sampai ketua umum urus besar hmi m fachruddin temu pimpin fraksi utus daerah mpr gedung dprmpr senayan jakarta selatan selasa 155</i>	22
4	<i>fachruddin tegang politik jadi pusat kerap korban tokoh tokoh daerah utama daerah sedang golak</i>	14
5	<i>jumlah tokoh jadi korban keras rektor institut agama islam negeri arraniry aceh syofwan idris tengku nashiruddin daud anggota dpr asal aceh jafar sidik hamzah ketua international forum for aceh</i>	29

Berdasarkan tabel diatas, jumlah kata dalam kalimat 5 merupakan terbanyak pada dokumen tersebut. kalimat 3 ini nantinya akan dijadikan acuan untuk menentukan Panjang kalimat pada masing-masing kalimat pada dokumen tersebut. untuk perhitungannya dapat dilihat dibawah ini.

$$F1(S_1) = \frac{13}{29} = 0,44$$

$$F1(S_2) = \frac{9}{29} = 0,31$$

$$F1(S_3) = \frac{22}{29} = 0,75$$

$$F1(S_4) = \frac{14}{29} = 0,48$$

$$F1(S_5) = \frac{29}{29} = 1$$

### 3.4.2 Posisi Kalimat (F2)

Fitur ini digunakan karena biasanya kalimat di awal dokumen memiliki skor lebih tinggi daripada di posisi terakhir (Rainarli & Dewi, 2018).

$$F2 = \frac{m-i}{m} \quad (3.2)$$

Disini  $m$  adalah total jumlah kalimat pada dokumen,  $i$  adalah posisi kalimat ke- $i$ .

$$F2(S_1) = \frac{5}{5} = 1$$

$$F2(S_2) = \frac{4}{5} = 0,8$$

$$F2(S_3) = \frac{3}{5} = 0,6$$

$$F2(S_4) = \frac{2}{5} = 0,4$$

$$F2(S_5) = \frac{1}{5} = 0,2$$

### 3.4.3 Kalimat Mengandung Data Numerik (F3)

Untuk meringkas teks perlu mempertimbangkan data numerik dalam dokumen, itu karena biasanya kalimat dengan itu memberikan informasi penting (Aulia *et al.*, 2021).

$$F3 = \frac{\# \text{ data numerik dalam kalimat}}{\# \text{ kata dalam kalimat}} \quad (3.3)$$

Berdasarkan pada Tabel 3. 5 persamaan untuk fitur kalimat mengandung data numerik dapat dilihat dibawah ini

$$F3(S_1) = \frac{0}{13} = 0$$

$$F3(S_2) = \frac{0}{9} = 0$$

$$F3(S_3) = \frac{1}{22} = 0,04$$

$$F3(S_4) = \frac{0}{14} = 0$$

$$F3(S_5) = \frac{0}{29} = 0$$

### 3.4.4 TF-ISF (F4)

Nilai TF-ISF meningkat seiring dengan bertambahnya jumlah kemunculan term dalam kalimat tetapi diimbangi dengan frekuensi kemunculan term dalam keseluruhan dokumen. (Mandal *et al.*, 2021). Rata-rata nilai TF-ISF yang diperoleh untuk suku-suku kalimat diambil untuk menghitung skor TF-ISF untuk kalimat tersebut. nilai tersebut dapat dihitung menggunakan

$$F4 = tf \times \left( \log \frac{n}{sf(t)} + 1 \right) \quad (3.4)$$

$$tf = \frac{\text{jumlah kemunculan term dalam kalimat}}{\text{total jumlah kata dalam kalimat}} \quad (3.5)$$

$$sf(t) = \frac{\text{jumlah kalimat yang mengandung kata tersebut}}{\text{total jumlah kalimat}} \quad (3.6)$$

$$isf = \log \frac{n}{sf(t)} + 1 \quad (3.7)$$

Di mana 'tf' adalah frekuensi istilah, 'n' adalah jumlah kalimat dalam dokumen dan 'sf(t)' adalah jumlah kalimat dalam dokumen yang mengandung istilah 't'.

Pada Tabel 3. 5 dokumen tersebut sudah dibagi menjadi kalimat-kalimat yang berjumlah 5. Untuk melakukan perhitungan TF-ISF kata “daerah” pada dokumen tersebut, pertama kali adalah menghitung *term frequency* (TF) dalam setiap kalimat. Persamaannya adalah sebagai berikut.

$$tf = \frac{\text{jumlah kemunculan term dalam kalimat}}{\text{total jumlah kata dalam kalimat}} \quad (3.8)$$

$$tf(S_1) = \frac{1}{13} = 0,07$$

$$tf(S_2) = \frac{1}{9} = 0,11$$

$$tf(S_3) = \frac{1}{22} = 0,04$$

$$tf(S_4) = \frac{2}{14} = 0,14$$

$$tf(S_5) = 0$$

Setelah itu menghitung Inverse *Sentence* Frequency (ISF) dari kata "daerah" dengan persamaan 3.7. Persamaan ini memperhitungkan frekuensi relatif kemunculan kata tersebut dalam kalimat-kalimat yang ada dalam suatu korpus teks. ISF digunakan untuk menilai seberapa informatif suatu kata dalam suatu konteks tertentu.

$$TF - ISF = tf * isf \quad (3.9)$$

Setelah mendapatkan nilai ISF, Langkah selanjutnya adalah menghitung nilai TF-ISF dengan persamaan berikut.

$$TF - ISF_1 = 0,07 * 1,096 = 0,076$$

$$TF - ISF_2 = 0,11 * 1,096 = 0,126$$

$$TF - ISF_3 = 0,04 * 1,096 = 0,043$$

$$TF - ISF_4 = 0,14 * 1,096 = 0,153$$

$$TF - ISF_5 = 0 * 1,096 = 0$$

Kemudian setelah melakukan perhitungan TF-ISF untuk semua kata dalam kalimat dan semua kalimat dalam dokumen, hasil akhir dari skor TF-ISF adalah

Tabel 3. 6 Skor TF-ISF

$S_{(n)}$	Skor TF-ISF
1	0.12942124329628701
2	0.1802368993210959
3	0.07411609528320763
4	0.12441208625363791
5	0.06799125912373025

### 3.4.5 Kemiripan Antar Kalimat (F5)

Fitur ini dihitung antara kalimat dan centroid. Centroid pada fitur ini diperoleh dari skor TF-ISF yang paling besar. (Shirwandkar & Kulkarni, 2018).

$$F5(S_i) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\text{Cos}(\text{sentence}_i \cdot \text{centroid})}{\|\text{sentence}_i\| \cdot \|\text{centroid}\|} \quad (3.10)$$

Penerapan pada fitur ini dapat dilihat pada Tabel 3. 7 berdasarkan persamaan untuk fitur diatas, yang menjadi centroid adalah kalimat ke 2 karena



memiliki skor TF-ISF terbesar. Kemudian setiap *vector* TF-ISF akan dibandingkan dengan centroid untuk menghasilkan nilai kemiripan antar kalimat.

Tabel 3. 7 Nilai Kemiripan Antar Kalimat

$S_{(n)}$	Hasil nilai kemiripan antar kalimat
1	0.23366709318314471
2	1.0
3	0.03495287740184593
4	0.40646286862852493
5	0.1345851696005682

Kalimat 2 memiliki nilai kemiripan 1 karena kalimat tersebut di similarity dengan kalimat itu sendiri. hasil semua ekstraksi fitur pada setiap kalimat dan juga labelnya dapat dilihat pada tabel dibawah ini

Tabel 3. 8 Hasil *Sentence Scoring*

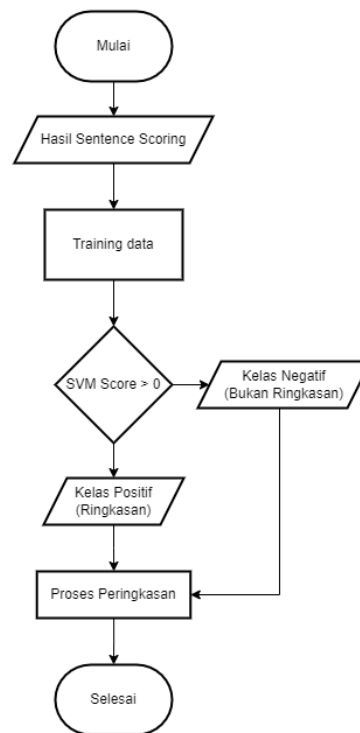
$S_{(n)}$	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>F5</b>	<b>Label</b>
1	0,44	1,0	0	0,12	0,23	1
2	0,31	0,8	0	0,18	1,0	1
3	0,75	0,6	0,04	0,07	0,03	1
4	0,48	0,4	0	0,12	0,4	-1
5	1,0	0,2	0	0,06	0,13	-1

Setelah proses ekstraksi fitur pada setiap kalimat untuk mendapatkan *Sentence Scoring* sudah selesai dilakukan, Tabel 3. 8 tersebut akan dijadikan sebagai inputan pada SVM sebagai vektor fitur.

### 3.5 Implementasi *Support Vector Machine*

Dalam penelitian ini, SVM model yang digunakan adalah binary classifier, karena didalam sistem hanya memisahkan kalimat pada dokumen kedalam kelas positif dan kelas negatif. Model SVM untuk klasifikasi biner memerlukan

pelabelan data pelatihan menggunakan variabel  $y_i = \pm 1$  (dengan tanda yang sesuai dengan kelas positif atau negatif dari kalimat dokumen pelatihan. (Winters-Hilt *et al.*, 2006). Alur yang digunakan untuk menentukan kelas hasil klasifikasi dengan menggunakan metode SVM digambarkan dalam sebuah *Flowchart* pada Gambar 3. 5



Gambar 3. 5 *Flowchart Support Vector Machine*

Hasil dari klasifikasi berdasarkan posisi nilai inputan pada *hyperplane*. Linear *hyperplane* digunakan untuk memisahkan kelas positif dan kelas negatif pada data *training*. Persamaannya adalah

$$w_i x_i + b = 0 \quad (3.11)$$

Dimana  $w$  adalah bobot pada masing-masing fitur yang berbentuk vektor  $n$ -dimensi,  $x_i$  adalah nilai parameter masing-masing fitur dan  $b$  adalah bias. Pada

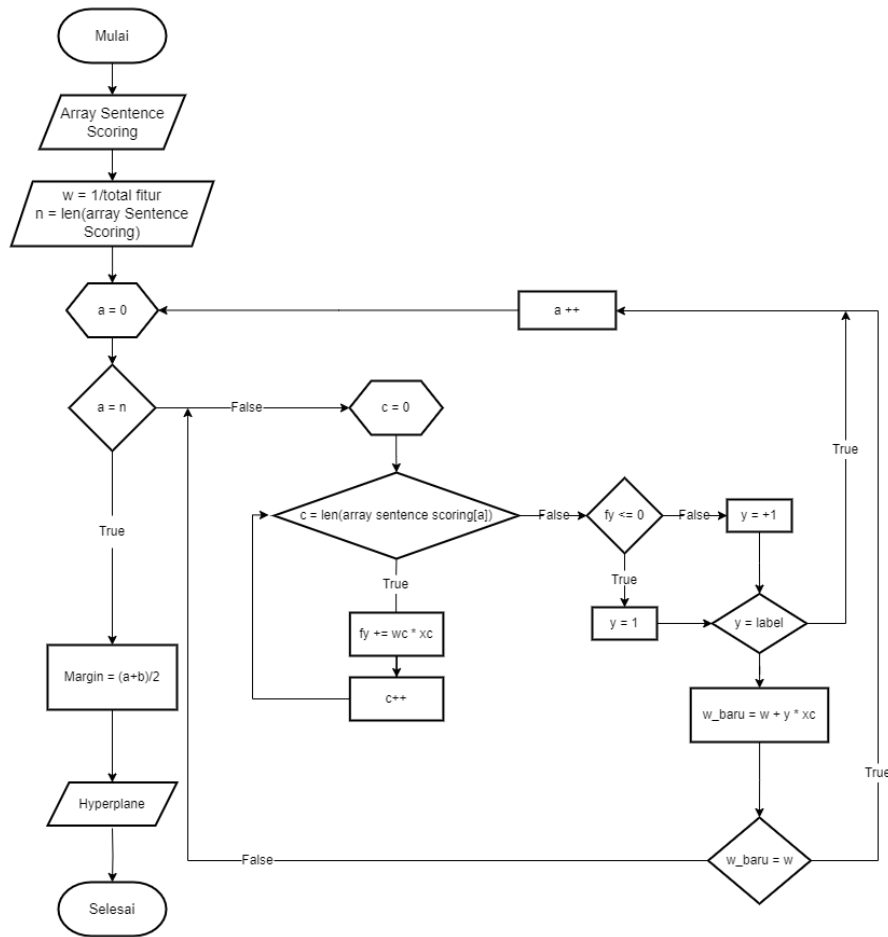
penelitian ini proses klasifikasi menggunakan fungsi kernel linear. Dari persamaan tersebut maka dihasilkan dua persamaan *hyperplane* untuk masing-masing kelas yaitu,

$$w_i x_i + b \begin{cases} \geq 1 & \text{untuk } y = +1 \\ \leq 1 & \text{untuk } y = -1 \end{cases} \quad (3.12)$$

Langkah berikutnya adalah menghitung parameter nilai dari setiap fitur yang ada. Fungsi ini akan digunakan sebagai bagian dari pembentukan *hyperplane*.

$$f(y) = \sum_{i=1}^n w_i x_i + b \quad (3.13)$$

Maka diperlukan untuk mencari nilai  $w$  dan  $b$  agar fungsi tersebut dapat diselesaikan. Algoritma klasifikasi Support Vector Machine (SVM) merupakan metode yang efektif untuk memisahkan kelas-kelas dalam data dengan menggunakan *hyperplane*. *Hyperplane* ini berperan sebagai batas pemisah antara kelas negatif dan positif dalam dataset. Untuk mengaplikasikan SVM pada data, diperlukan penentuan nilai dari parameter  $w$  (vektor bobot) dan  $b$  (bias). Proses ini dapat diformulasikan sebagai sebuah masalah optimasi, di mana SVM berusaha untuk mencari nilai  $w$  dan  $b$  yang meminimalkan fungsi objektif yang mencakup elemen-elemen dari vektor bobot dan faktor penalti untuk titik-titik yang berada di sisi yang salah dari *hyperplane*. Melalui proses pelatihan, SVM mampu menghasilkan *hyperplane* yang efektif memisahkan kelas-kelas dalam data dengan tepat, memungkinkan pengklasifikasiannya secara akurat terhadap data baru berdasarkan posisi terhadap *hyperplane* tersebut. *Flowchart* yang menjelaskan bagaimana *hyperplane* dibentuk dapat dilihat pada Gambar 3. 6



Gambar 3. 6 Flowchart hyperplane

Langkah awal yang dilakukan untuk mendapatkan *hyperplane* optimal adalah dengan menginisialisasi bobot awal. Pada penelitian ini bobot awal dari hasil perhitungan untuk mencari *hyperplane* adalah

$$w_i^0 = \frac{1}{n} \quad (3.14)$$

Dimana  $w_i^0$  adalah inisialisasi bobot awal dan  $n$  adalah jumlah fitur yang terdapat pada proses *Sentence Scoring*. Berikut nilai  $w_i^0$  yang diperoleh dengan menggunakan persamaan 3.14

$$w_i^0 = \frac{1}{5} = 0,2$$

Selanjutnya adalah menentukan nilai bias yang diperoleh dari persamaan 3.15

$$b^0 = \frac{1}{n+1} \quad (3.15)$$

$$b^0 = \frac{1}{(5+1)} = 0,33$$

Dari nilai  $f(y)$  yang dihasilkan dari persamaan 3.13, telah ditentukan jika  $f(y) < 0$  maka  $y_i = -1$  dan sebaliknya, jika nilai  $f(y) > 0$  maka  $y_i = +1$ . Sebagai contoh, disini akan menggunakan  $S_1$  pada tabel 3.9.

$$f(y) = ((0,2 \times 0,44) + (0,2 \times 1,0) + (0,2 \times 0) + (0,2 \times 0,12) + (0,2 \times 0,23)) + 0,33$$

$$f(y) = (0,088 + 0,2 + 0 + 0,024 + 0,046 + 0,33)$$

$$f(y) = 0,688$$

$$f(y) > 0, \text{ maka } y_i = +1$$

Setelah melakukan perhitungan untuk menentukan nilai parameter, ditemukan bahwa nilai adalah  $y = +1$ , sehingga dapat disimpulkan bahwa data  $S_1$  tergolong ke dalam kelas positif. Jika nilai label dari sebuah data, yaitu  $y_i$ , tidak sama dengan hasil klasifikasi dari model pada data tersebut, maka diperlukan perbaruan pada bobot. Bobot baru akan dihitung menggunakan persamaan dibawah ini

$$w_{baru} = w_{lama} + y_i \cdot x_{ij} \quad (3.16)$$

Disini  $w_{baru}$  adalah bobot baru yang akan dicari nilainya,  $w_{lama}$  adalah bobot awal yang sudah diinisialisasi pada persamaan 3.14  $y_i$  adalah label yang ada pada data pelatihan dan  $x_{ij}$  adalah nilai hasil dari perhitungan *hyperplane* sebelumnya.

Langkah berikutnya adalah menerapkan sistem yang telah dibangun pada data pelatihan dan data pengujian. Saat diterapkan pada data pelatihan, sistem akan menghasilkan model klasifikasi SVM yang kemudian akan digunakan untuk menguji data pengujian sebagai input, dengan menggunakan nilai bobot dan bias pada fitur yang sama dengan data pelatihan. Implementasi sistem untuk menentukan kelas dari data pengujian dilakukan dengan memperhitungkan nilai  $f(y)$  yang dihasilkan dari persamaan 3.17. Jika nilai  $f(y) \leq 0$ , maka kalimat tersebut diklasifikasikan sebagai kelas negatif dan bukan sebagai ringkasan, sebaliknya jika nilai  $f(y) \geq 0$ , maka termasuk ke dalam kelas positif dan merupakan kalimat yang masuk dalam peringkasan.

### 3.6 Proses Peringkasan

Klasifikasi Support Vector Machine (SVM) adalah sebuah metode dalam pembelajaran mesin yang bertujuan untuk memisahkan data ke dalam kategori-kategori yang telah ditentukan. Dalam konteks aplikasinya pada proses ringkasan teks, SVM dapat digunakan untuk mengklasifikasikan setiap kalimat ke dalam dua kategori, yaitu label akhir 1 dan -1. Label akhir 1 menandakan bahwa kalimat tersebut memiliki tingkat relevansi yang tinggi dan layak untuk dimasukkan ke dalam ringkasan, sementara label -1 menunjukkan bahwa kalimat tersebut kurang relevan atau bahkan tidak relevan sama sekali untuk dimasukkan ke dalam ringkasan.

Setelah dilakukan proses klasifikasi SVM, langkah selanjutnya adalah menyusun ringkasan berdasarkan hasil klasifikasi tersebut. Kalimat-kalimat yang mendapatkan label akhir 1 akan dimasukkan ke dalam ringkasan sebagai bagian yang signifikan dari dokumen asli. Sebaliknya, kalimat-kalimat yang mendapatkan label -1 akan dianggap kurang relevan dan dibuang dari ringkasan. Dengan demikian, proses ini membantu menyaring dan menyusun informasi yang paling penting dan relevan untuk disajikan dalam sebuah ringkasan yang singkat dan informatif.

Dengan menggunakan SVM dalam proses klasifikasi, penggunaan teknik ini tidak hanya memudahkan identifikasi kalimat-kalimat yang memiliki nilai tambah untuk ringkasan, tetapi juga meningkatkan efisiensi dan ketepatan dalam menyusun ringkasan teks. Dengan demikian, penggunaan SVM memberikan kontribusi penting dalam pengembangan sistem ringkasan otomatis yang dapat mengolah dan menyajikan informasi secara lebih efektif.

## BAB IV

### UJI COBA DAN PEMBAHASAN

#### 4.1 Dataset

Untuk membentuk dataset baru, langkah awalnya melibatkan pengambilan artikel dari dataset INDOSUM yang terdiri dari 14.000 berita *online*, sebagaimana terlihat pada Gambar 3.1. Proses pertama melibatkan pencarian berita dengan menggunakan query tertentu yang akan membentuk list story. Istilah "story" di sini merujuk pada kumpulan berita *online* yang berkesinambungan dalam satu topik atau pembahasan tertentu, dengan kata kunci query yang telah disusun seperti pada Tabel 4. 1. Dengan demikian, langkah-langkah ini membentuk dasar pembentukan dataset baru yang lebih spesifik dan terfokus pada topik atau pembahasan tertentu sesuai dengan karakteristik query yang telah ditentukan sebelumnya.

Tabel 4. 1 Query Pembentukan Story multi dokumen

No	Query	Judul
1	['presiden', 'perang', 'melawan']	['Mati Dalam Operasi Antinarkotika Filipina Pro', 'Perang Duterte Lawan Narkoba Dipergencar', 'Filipina Diserang Isu Pelanggaran Ham', 'Rodrigo Duterte Diselidiki Mahkamah Internasional']
2	['laptop', 'gaming', 'asus']	['Asus Xiu Laptop Multimedia Gaming Ready', 'Asus Rilis Laptop Gim Seharga Motor Sport', 'Asus Ramu Notebook Rog Gvik']
3	['gaming', 'acer']	['Asia Pacific Predator League Resmi Dimulai', 'Acer Nitro Spin Laptop Gaming In Terbaru', 'Philips Sedang Garap Monitor K']
...	...	...
48	['musisi', 'synchronize', 'festival']	['Musisi Pengisi Synchronize Festival Siap Beri', 'Musik Dangdut Masuk Dalam Synchronize Fest ', 'Jimi The Upstairs Berharap Tampil Di Synchronize F', 'Pujian Jokowi Untuk Musisi Yang Tampil Di Synchron']
49	['konser', 'jakarta', 'celine', 'dion']	['Celine Dion Akan Gelar Konser Pertama Di Indonesia', 'Netizen Protes Tiket Konser Dp Mobil Celine Dion', 'Celine Dion Ogah Beri Saran Lady Gaga Soal Konser ']
50	['film', 'marvel', '2017']	['Lima Hal Yang Harus Diketahui Sebelum Menonton Tho', 'Spider Man Kalahkan Empat Tim Avengers', 'Chris Hemsworth Nyaris Menolak Peran Thor']



Setelah memperoleh story multi dokumen, langkah selanjutnya adalah melakukan pelabelan terhadap setiap kalimat dalam story tersebut oleh lima orang responden. Proses pelabelan ini dilakukan untuk memilih kalimat-kalimat yang akan dimasukkan ke dalam ringkasan dan yang tidak. Setiap responden memberikan label pada setiap kalimat berdasarkan kepentingan dan relevansinya dalam meringkas story. Data dari responden tersebut kemudian akan digunakan sebagai ringkasan manusia, yang selanjutnya akan dibandingkan dengan ringkasan yang dihasilkan oleh sistem. Berikut ini adalah data story multi dokumen yang sudah dikumpulkan sebanyak 50 story.

id_story	id_document	query	kategori	judul	content	responden1	responden2	responden3	responden4	responden5	label
40	41	[3368, 10184, 13242]	[juara, euro, 2016, portugal, prancis]	olahraga	[Portugal Sebut Juara Euro Bukan Kebenarannya...]	[[Merdeka.com, -, Pelatih, Portugal, -, Ferna...]]	[[True, True, True, False, True, False, True, False, ...]]	[[False, False, False, False, False, False, False, ...]]	[[True, True, True, False, True, False, False, False, ...]]	[[True, True, True, False, True, False, False, False, ...]]	[[True, True, True, False, True, False, False, False, ...]]
41	42	[6930, 11937]	[asian, games, indonesia, jokow]	olahraga	[Indonesia Buka Peluang Jadi Tuan Rumah Olimpi...]	[[JUARA.NET, -, JAKARTA, -, Presiden, Joko, W...]]	[[True, True, True, False, True, False, True, True, ...]]	[[True, True, True, False, True, False, True, True, ...]]	[[True, True, True, False, True, False, True, True, ...]]	[[True, True, True, False, True, False, True, True, ...]]	[[True, True, True, False, True, False, True, True, ...]]
42	43	[7246, 10326]	[olimpiade, 2016, indonesia, medali, emas]	olahraga	[Dani Alves Jadi Kapten Timnas Brasil, Tontow...]	[[Jakarta, -, CNN, Indonesia, -, -, Neymar, L...]]	[[False, False, True, True, True, True, True, True, ...]]	[[False, False, True, True, True, True, True, True, ...]]	[[False, False, True, True, True, True, True, True, ...]]	[[False, False, True, True, True, True, True, True, ...]]	[[True, False, False, False, False, False, False, False, ...]]
43	44	[3881, 4175, 11167]	[film, marvel, 2017]	showbiz	[Lima Hal Yang Harus Diketahui Sebelum Menon...]	[[lilah, lima, hal, menarik, yang, harus, An...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]
44	45	[5580, 14049]	[liris, album, taylor, swift]	showbiz	[Taylor Swift Membuat Aplikasi Jejaring Sosial...]	[[Jakarta, -, CNN, Indonesia, -, -, Jejang, F...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]
45	46	[3507, 3854, 10428]	[afgan, album]	showbiz	[Agniez Mo Slap Tampilkan Lagu Terbaru, Lagu Du...]	[[Jakarta, (, ANTARA, News, ), -, , Penyanyi...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]
46	47	[280, 3792, 7234]	[konser, jakarta, celine, dion]	showbiz	[Celine Dion Akan Gelar Konser Pertama Di Indo...]	[[Jakarta, -, CNN, Indonesia, -, -, Penyanyi...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]
47	48	[1678, 14065]	[tiket, konser, ed, sheeran, jakarta]	showbiz	[Tiket Ed Sheeran Ludes Terjual Dalam Sehari, ...]	[[Jakarta, -, CNN, Indonesia, -, -, Penjualan...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]
48	49	[841, 5836, 6810, 7814]	[musisi, synchronize, festival]	showbiz	[Musisi Pengisi Synchronize Festival Siap Ber...]	[[Jakarta, -, CNN, Indonesia, -, -, Beberapa...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]
49	50	[904, 13834]	[resepsi, pernikahan, kahiyang, putri, presiden]	showbiz	[Raisa Dan Ashanti Iku Kirim Doa Ke Kahiyang ...]	[[Solo, -, CNN, Indonesia, -, -, Penyanyi, Ra...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]	[[True, True, True, True, True, True, True, True, ...]]

Gambar 4. 1. Dataset Story Multi Dokumen

Gambar 4.1 menunjukkan bahwa penelitian ini menggunakan sebanyak 50 berita multi-dokumen yang telah dikelompokkan. Setiap berita multi-dokumen terbentuk dari rentang 3 hingga 5 berita single dokumen. Untuk meningkatkan validitas penelitian, proses pelabelan dilakukan kembali dengan melibatkan

partisipasi dari 5 responden. Dalam tahap ini, setiap kalimat yang terdapat dalam berita multi-dokumen dinilai sebagai "*True*" atau "*False*." Penilaian ini bertujuan untuk menentukan apakah suatu kalimat seharusnya dimasukkan ke dalam ringkasan atau tidak. Label akhir yang dihasilkan adalah label "*True*," yang diberikan kepada kalimat yang memperoleh lebih dari 50% dari total bobot jawaban yang terkumpul dari responden. Contoh ilustrasi dari proses pelabelan setiap kalimat dapat ditemukan pada Tabel 4. 2 yang terlampir di bawah ini. Proses ini dibuat dengan cermat untuk memastikan bahwa penentuan keberadaan suatu kalimat dalam ringkasan didasarkan pada penilaian manusia yang akurat dan dapat dipertanggungjawabkan.

Tabel 4. 2 Contoh proses pelabelan

Kalimat	Annotator					Sum	label ringkasan
	1	2	3	4	5		
Kalimat 1	<i>True</i>	<i>False</i>	<i>True</i>	<i>True</i>	<i>True</i>	4	<i>True</i>
Kalimat 2	<i>False</i>	<i>True</i>	<i>False</i>	<i>False</i>	<i>False</i>	1	<i>False</i>
Kalimat 3	<i>True</i>	<i>True</i>	<i>False</i>	<i>False</i>	<i>False</i>	2	<i>False</i>
Kalimat 4	<i>True</i>	<i>True</i>	<i>True</i>	<i>False</i>	<i>False</i>	3	<i>True</i>
Kalimat 5	<i>True</i>	<i>False</i>	<i>False</i>	<i>False</i>	<i>False</i>	1	<i>False</i>
...	...	...	...	...	...	...	...
Kalimat 100	<i>True</i>	<i>True</i>	<i>False</i>	<i>False</i>	<i>True</i>	3	<i>True</i>

Langkah berikutnya dalam penelitian ini adalah melakukan proses *Sentence Scoring* pada setiap berita multi dokumen. Proses ini melibatkan ekstraksi fitur pada setiap kalimat dengan menggunakan lima fitur, yaitu panjang kalimat, posisi kalimat, data numerik, TF-ISF (*Term Frequency-Inverse Sentence Frequency*), dan kemiripan antar kalimat. Setelah ekstraksi fitur, dilakukan proses klasifikasi menggunakan metode *Support Vector Machine* (SVM) untuk memisahkan kalimat-kalimat yang dianggap penting dan tidak penting berdasarkan label yang telah

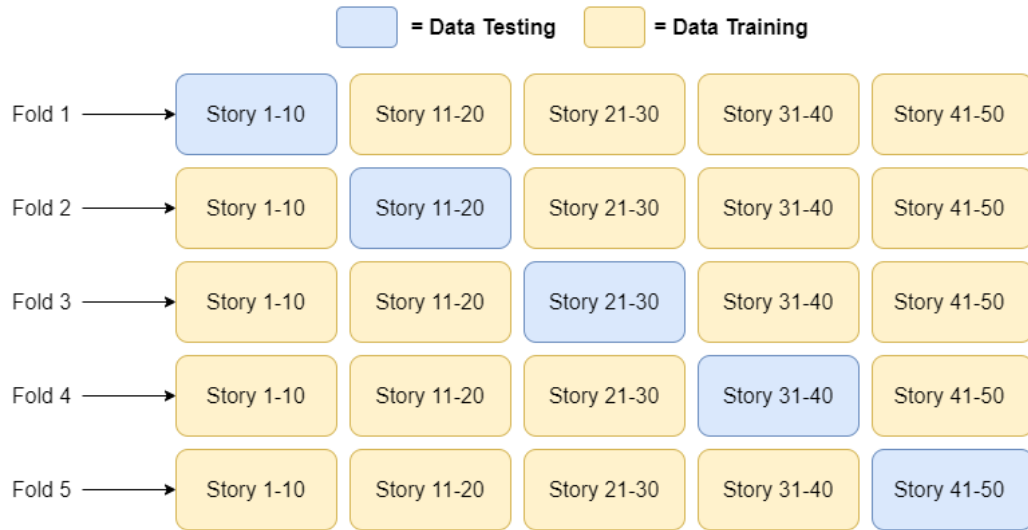
diberikan pada setiap kalimat. Hasil dari proses *Sentence Scoring* yang terdapat pada Tabel 4. 3 akan digunakan sebagai input untuk klasifikasi SVM. Label pada data diubah dari *True* menjadi 1 dan *False* menjadi -1. Nilai 1 menunjukkan kriteria kalimat ringkasan dan nilai -1 mewakili kriteria kalimat bukan ringkasan. Perubahan status menjadi numerik bertujuan untuk mempermudah proses klasifikasi oleh SVM. Langkah-langkah ini untuk meningkatkan efisiensi dan ketepatan dalam menentukan relevansi serta signifikansi kalimat-kalimat dalam konteks berita multi dokumen.

Tabel 4. 3. *Sentence Scoring*

<b>Id</b>	<b>Id_story</b>	<b>Label</b>	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>F5</b>
0	0_0_0	1	1.000000	1.000000	0.000000	0.715949	0.054589
1	0_0_1	1	0.882353	0.833333	0.133333	0.737574	0.017825
2	0_0_2	-1	0.588235	0.666667	0.100000	0.718869	0.000000
3	0_0_3	1	0.117647	0.500000	0.000000	0.397940	0.089861
4	0_0_4	-1	0.882353	0.333333	0.133333	0.799085	1.000000
...	...	...	...	...	...	...	...
788	9_3_13	1	1.000000	0.277778	0.100000	1.106109	0.000000
789	9_3_14	1	0.850000	0.222222	0.117647	1.262356	0.000000
790	9_3_15	-1	0.150000	0.166667	0.000000	0.908637	0.000000
791	9_3_16	-1	0.300000	0.111111	0.000000	0.994825	0.000000
792	9_3_17	-1	0.800000	0.055556	0.000000	1.165182	0.000000

## 4.2 Skenario Uji Coba

Dalam melakukan uji coba menggunakan SVM, data *training* dan data *testing* dibagi dengan menggunakan teknik *K Fold Cross Validation*. Skenario uji coba dilakukan dari *Fold 1* sampai dengan *Fold 5* untuk mengukur nilai ROUGE-1 pada masing-masing story dalam data set. Masing-masing *Fold* berisikan 10 story sebagai data *testing* dan 40 story lainnya menjadi data *training*. Percobaan tersebut dilakukan sampai *Fold 5* untuk memperoleh nilai ROUGE-1 dari masing-masing story multi dokumen.



Gambar 4. 2. *K-Fold Cross Validation*

### 4.3 Hasil Uji Coba

Pada skenario uji coba yang telah dijelaskan pada sub bab 4.1 pengujian dilakukan untuk membandingkan ringkasan manusia dan ringkasan dari sistem. Pada skripsi ini menggunakan metrik ROUGE-1 dalam mendapatkan nilai *Recall*, *Precision*, dan *F-Measure*. ROUGE-1, dalam konteks evaluasi sistem ringkasan teks menggunakan metrik *Recall* dan *Precision*. *Recall* ROUGE-1 mengukur kemampuan sistem untuk menangkap informasi relevan dari ringkasan mesin (sistem). Di sisi lain, *Precision* ROUGE-1 mengukur seberapa banyak informasi relevan yang dimasukkan dalam ringkasan oleh sistem. Kombinasi dari *Recall* dan *Precision* memberikan gambaran keseluruhan tentang kemampuan sistem dalam menangkap dan menyajikan informasi yang sesuai dan relevan dalam ringkasannya. Dalam konteks peringkasan berita, *Recall* akan memberi tahu seberapa baik sistem

dapat mencakup semua informasi kunci yang ada dalam ringkasan manusia. Berikut ini adalah persamaan pada metrik ROUGE-1

$$Recall = \frac{\text{Jumlah kata yang sama pada ringkasan manusia dan sistem}}{\text{Jumlah kata pada ringkasan manusia}} \quad (4.1)$$

$$Presicion = \frac{\text{Jumlah kata yang sama pada ringkasan manusia dan sistem}}{\text{Jumlah kata pada ringkasan sistem}} \quad (4.2)$$

$$F - Measure = 2 \times \frac{(Recall \times Presicion)}{(Recall + Presicion)} \quad (4.3)$$

#### 4.3.1 Percobaan *Fold 1*

Pada *Fold 1*, story yang menjadi data *testing* dan yang menjadi ringkasan adalah story 1-10, sedangkan yang menjadi data *training* adalah story 11-50. Tabel 4. 4 menampilkan hasil statistik dokumen dari hasil peringkasan manusia dan peringkasan sistem untuk berita multi dokumen pada story 1-10

Tabel 4. 4 Statistik Jumlah Kata Dalam Ringkasan *Fold 1*

List Story	kata unik pada ringkasan manusia	kata unik pada ringkasan mesin	kata unik pada ringkasan mesin dan manusia
Story 1	491	683	411
Story 2	397	336	240
Story 3	287	532	183
Story 4	121	465	100
Story 5	118	316	98
Story 6	330	645	297
Story 7	222	465	222
Story 8	177	531	177
Story 9	434	1407	400
Story 10	362	780	322

Selanjutnya adalah menghitung nilai *Recall*, *Presicion*, dan *F1-Score* pada setiap story dengan persamaan 4.1, 4.2, dan 4.3. Hasil dari perhitungan ringkasan dari *Fold 1* ditunjukkan pada Tabel 4. 5

Tabel 4. 5 Hasil Perhitungan ROUGE-1 *Fold 1*

List Story	Recall	Precision	F-Measure
Story 1	0.837	0.601	0.700
Story 2	0.604	0.714	0.654
Story 3	0.658	0.343	0.451
Story 4	0.826	0.215	0.341
Story 5	0.830	0.310	0.451
Story 6	0.9	0.460	0.609
Story 7	1.0	0.477	0.646
Story 8	1.0	0.333	0.5
Story 9	0.921	0.284	0.434
Story 10	0.889	0.412	0.563

Berikut ini adalah contoh dari hasil ringkasan manusia dan sistem pada story 8 dengan query “tesla” , “mobil”, “X” dengan hasil nilai *Recall* 1.0 yang terdapat pada Gambar 4. 3 dan Gambar 4. 4

Hadirnya Tesla Model X menjawab janji Presiden Direktur Prestige Motorcars Rudy Salim ketika merilis Model S beberapa waktu lalu. Mereka pertama kali meluncurkan Tesla Model S pada tahun 2015 hingga hari ini mereka memiliki model terbaru dari sedan listrik itu yakni P100D. " Kita dari sisi pedagang (showroom) tentunya mempunyai varian mobil yang banyak dan semuanya secara comfort dan kemampuan dari sisi mobil sendiri belum ada yang mengalahkan Tesla. Tesla Model X merupakan crossover yang menggunakan fitur falcon wing untuk mengakses kursi baris kedua dan ketiga. Mobil ini adalah kelanjutan pengembangan dari platform sedan full - size mereka Model S. Tesla memproduksi mobil ini di Fremont, California, dan dikirim perdana ke konsumen pada September 2015. Soal fitur, mobil ini memiliki fitur keselamatan aktif meliputi collision avoidance dan automatic emergency braking, sistem all - wheel drive secara elektrik untuk memaksimalkan efisiensi dan traksi, headlamp LED, power - Folding serta side mirrors. Selain pintu, Tesla Model X juga punya kaca depan yang sangat lebar untuk memberikan pengalaman berkendara yang berbeda, kunci keyless entry, ruang lapang pada kabin dan bagasi depan yang bisa mengakomodasi dua stroller sekaligus. Langkah Tesla yang segera menjual Model 3 - - sedan kompak listrik murah - - nampaknya membuat BMW tertarik untuk menyiapkan pesaing. Menurut laporan media Jerman, Handelsblatt, BMW akan meluncurkan sebuah versi listrik dari model Seri 3 di Frankfurt Motor Show pada September. Sebagai model yang dipersiapkan sebagai penantang Tesla Model 3, BMW Seri 3 listrik mengemas baterai yang memiliki daya jelajah yang lebih baik. Kabarnya, sedan kompak listrik BMW menjanjikan jarak jelajah 400 km. Selain full listrik, BMW kabarnya turut menyediakan opsi plug-in hybrid, yang bila mengandalkan modus listrik bisa dibuat berkendara sejauh 50 kilometer. Nah, kalo kita berbicara Tesla Model 3, mengacu pada bocoran, mobil itu memiliki fitur infotainment yang dipusatkan pada layar berukuran besar dengan bentang 15 inci. Kabarnya, Tesla Model 3 akan dipasarkan dalam dua pilihan baterai ; 55 kWh and 70 kWh, yang dijual dengan harga 35 ribu dolar Amerika (Rp 457 juta) dan 42 ribu dolar Amerika (Rp 548 juta). Dibanding Model S atau Model X, Model 3 jauh lebih menarik bagi sebagian besar konsumen karena harganya yang sangat terjangkau – terutama untuk kategori mobil elektrik – yakni \$ 35.000 untuk konfigurasi paling mendasarnya yang sudah bisa menempuh jarak 346 kilometer dalam satu kali charge. Model 3 dijadwalkan bakal masuk tahap produksi mulai tahun ini, dan Tesla rupanya sudah mulai mengerjakan 30 unit pertamanya sampai selesai. 30 hanyalah awal, karena jumlahnya akan terus naik secara eksponensial : 100 unit di bulan Agustus, di atas 1.500 unit di bulan September, dan pada bulan Desember Tesla menargetkan bisa memproduksi sebanyak 20.000 Model 3 per bulannya. Masuk tahun 2018, Tesla berharap bisa memproduksi 10.000 unit Model 3 setiap minggu. Performanya pun sekelas mobil sport, dengan akselerasi 0 – 100 km / jam dalam waktu 5,6 detik – meski masih kalah jauh dari Model S.

Gambar 4. 3 Hasil Ringkasan Manusia Story 8

Menurut laporan media Jerman, Handelsblatt, BMW akan meluncurkan sebuah versi listrik dari model Seri 3 di Frankfurt Motor Show pada September. Kabarnya, Tesla Model 3 akan dipasarkan dalam dua pilihan baterai ; 55 kWh and 70 kWh, yang dijual dengan harga 35 ribu dolar Amerika (Rp 457 juta) dan 42 ribu dolar Amerika (Rp 548 juta). Dibanding Model S atau Model X, Model 3 jauh lebih menarik bagi sebagian besar konsumen karena harganya yang sangat terjangkau – terutama untuk kategori mobil elektrik – yakni \$ 35.000 untuk konfigurasi paling mendasarnya yang sudah bisa menempuh jarak 346 kilometer dalam satu kali charge. Model 3 dijadwalkan bakal masuk tahap produksi mulai tahun ini, dan Tesla rupanya sudah mulai mengerjakan 30 unit pertamanya sampai selesai. 30 hanyalah awal, karena jumlahnya akan terus naik secara eksponensial : 100 unit di bulan Agustus, di atas 1.500 unit di bulan September, dan pada bulan Desember Tesla menargetkan bisa memproduksi sebanyak 20.000 Model 3 per bulannya.

Gambar 4. 4 Hasil Ringkasan Sistem Story 8

### 4.3.2 Percobaan *Fold 2*

Tabel 4. 6 menampilkan hasil statistik dokumen dari hasil peringkasan manusia dan peringkasan sistem untuk berita multi dokumen pada story 11-20.

Tabel 4. 6 Statistik Jumlah Kata Dalam Ringkasan *Fold 2*

List Story	kata unik pada ringkasan manusia	kata unik pada ringkasan mesin	kata unik pada ringkasan mesin dan manusia
Story 11	146	251	146
Story 12	401	459	343
Story 13	321	295	240
Story 14	212	414	163
Story 15	518	1165	457
Story 16	215	385	197
Story 17	172	335	146
Story 18	117	324	117
Story 19	713	933	628
Story 20	247	298	201

Selanjutnya adalah menghitung nilai *Recall*, *Presicion*, dan *F1-Score* pada setiap story dengan persamaan 4.1, 4.2, dan 4.3. Kemudian hasil dari perhitungan ringkasan pada *Fold 2* ditunjukkan pada Tabel 4. 7 dibawah ini.

Tabel 4. 7 Hasil Perhitungan ROUGE-1 *Fold 2*

List Story	Recall	Precision	F-Measure
Story 11	1.0	0.581	0.735
Story 12	0.855	0.747	0.797
Story 13	0.747	0.813	0.779
Story 14	0.768	0.393	0.520
Story 15	0.882	0.392	0.543
Story 16	0.916	0.511	0.656
Story 17	0.848	0.435	0.575
Story 18	1.0	0.361	0.530
Story 19	0.880	0.673	0.763
Story 20	0.813	0.674	0.737

Berikut ini adalah contoh dari hasil ringkasan pada story 11 dengan query “korban” , “longsor”, “ponorogo” dengan hasil nilai *Recall* 1.0 yang terdapat pada Gambar 4. 5 dan Gambar 4. 6

Rimanews - Badan Nasional Penanggulangan Bencana (BNPB) menyatakan pencarian korban longsor di Ponorogo, Jawa Timur, terkendala medan bencana yang sulit, salah satunya karena tingginya longsor tanah yang menutupi area permukiman. " Tebal tanah mencapai 20 meter, lebar 200 meter dan panjang longsor dari mahkota longsor hingga titik terakhir hingga dua kilometer, " kata Kepala Pusat Data, Informasi dan Humas BNPB Sutopo Purwo Nugroho dalam jumpa persnya di Jakarta, hari ini. Sutopo menjelaskan alat-alat berat sudah didatangkan guna mempermudah proses pencarian korban, namun medan yang sulit membuat upaya pencarian korban berjalan lebih lambat dari seharusnya. Tingginya timbunan tanah akibat longsor di Ponorogo, menurut Sutopo, membuat proses pencarian korban tidak bisa secepat ketika bencana longsor di Banjarnegara, Jawa Tengah. Wakil Gubernur Jawa Timur, Saifullah Yusuf, mengatakan 27 orang korban tanah longsor di Dukuh Tangkil, Desa Pulung, Kecamatan Banaran, Kabupaten Ponorogo, belum ditemukan. " Data yang masuk sampai Sabtu pukul 19.00 WIB atau bersamaan dihentikannya sementara pencarian, masih 27 orang yang diperkirakan tertimbun tanah, " ujarnya kepada wartawan di Surabaya, Sabtu (1 / 4). " Memang awalnya ada yang mengatakan 17 orang, 28 orang, namun laporan dari Bupati Ponorogo Ipong Muchlissoni yang berada di lapangan, sebanyak 27 orang yang dinyatakan hilang, " ucapnya. Pemprov Jatim juga telah menyiagakan logistik dan membuka dapur umum 24 jam untuk pengungsi dan korban longsor.

Gambar 4. 5 Hasil Ringkasan Manusia Story 11

Rimanews - Badan Nasional Penanggulangan Bencana (BNPB) menyatakan pencarian korban longsor di Ponorogo, Jawa Timur, terkendala medan bencana yang sulit, salah satunya karena tingginya longsor tanah yang menutupi area permukiman. " Tebal tanah mencapai 20 meter, lebar 200 meter dan panjang longsor dari mahkota longsor hingga titik terakhir hingga dua kilometer, " kata Kepala Pusat Data, Informasi dan Humas BNPB Sutopo Purwo Nugroho dalam jumpa persnya di Jakarta, hari ini. Wakil Gubernur Jawa Timur, Saifullah Yusuf, mengatakan 27 orang korban tanah longsor di Dukuh Tangkil, Desa Pulung, Kecamatan Banaran, Kabupaten Ponorogo, belum ditemukan. " Data yang masuk sampai Sabtu pukul 19.00 WIB atau bersamaan dihentikannya sementara pencarian, masih 27 orang yang diperkirakan tertimbun tanah, " ujarnya kepada wartawan di Surabaya, Sabtu (1 / 4).

Gambar 4. 6 Hasil Ringkasan Sistem Story 11



### 4.3.3 Percobaan *Fold 3*

Tabel 4. 8 menampilkan hasil statistik dokumen dari hasil peringkasan manusia dan peringkasan sistem untuk berita multi dokumen pada story 21-30.

Tabel 4. 8 Statistik Jumlah Kata Dalam Ringkasan *Fold 3*

List Story	kata unik pada ringkasan manusia	kata unik pada ringkasan mesin	kata unik pada ringkasan mesin dan manusia
Story 21	219	424	204
Story 22	495	693	438
Story 23	611	1155	610
Story 24	517	702	434
Story 25	417	670	364
Story 26	345	854	344
Story 27	270	540	270
Story 28	139	432	139
Story 29	339	967	337
Story 30	429	807	409

Selanjutnya adalah menghitung nilai *Recall*, *Presicion*, dan *F1-Score* pada setiap story dengan persamaan 4.1, 4.2, dan 4.3. Kemudian hasil dari perhitungan ringkasan pada *Fold 3* ditunjukkan pada Tabel 4. 9 dibawah ini

Tabel 4. 9 Hasil Perhitungan ROUGE-1 *Fold 3*

List Story	<i>Recall</i>	<i>Presicion</i>	<i>F-Measure</i>
Story 21	0.931	0.481	0.634
Story 22	0.884	0.632	0.737
Story 23	0.998	0.528	0.690
Story 24	0.839	0.618	0.712
Story 25	0.872	0.543	0.669
Story 26	0.997	0.402	0.573
Story 27	1.0	0.5	0.666
Story 28	1.0	0.321	0.486
Story 29	0.994	0.348	0.516
Story 30	0.953	0.506	0.661

Berikut ini adalah hasil peringkasan dari story 27 dengan query “erupsi”, “gunung”, “sinabung” dengan hasil nilai *Recall* 1.0 yang terdapat pada Gambar 4. 7 dan Gambar 4. 8

Gunung Sinabung kembali mengalami erupsi. Menurut Kepala Humas BNPB Sutopo Purwonugroho, Sabtu (20 / 5), erupsi terjadi pada pukul 06.46 WIB. Sutopo mengimbau agar masyarakat dan pengunjung tidak melakukan aktivitas di dalam radius 3 Km dari puncak, dan dalam jarak 7 Km untuk sektor Selatan - Tenggara, di dalam jarak 6 km untuk sektor Tenggara - Timur, serta di dalam jarak 4 km untuk sektor Utara - Timur G. Sinabung. Mengingat telah terbentuk bendungan alam di hulu Sungai Laborus, maka penduduk yang bermukim dan beraktivitas di sekitar hilir daerah aliran sungai Laborus agar tetap menjaga kewaspadaan karena bendungan ini sewaktu-waktu dapat jebol, bila tidak kuat menahan volume air sehingga mengakibatkan lahar / banjir bandang ke hilir. " BPBD Kabupaten Tanah Karo segera melakukan sosialisasi ancaman bencana lahar / banjir bandang ini ke penduduk yang bermukim dan beraktivitas di sepanjang hilir dan sekitar Sungai Laborus, " tutupnya. Suara.com - Gunung Sinabung di Sumatera Utara erupsi untuk kesekian kalinya, pada Senin (19 / 2 / 2018). Menanggapi kejadian itu, Ketua DPR Bambang Soesatyo meminta Badan Nasional Penanggulangan Bencana pusat, berkoodinasi dengan BNPB daerah untuk memperbaiki sarana dan prasarana yang terkait dengan keselamatan dan kesehatan warga di Kaki Gunung Sinabung. " Kita meminta Pemerintah Daerah dan Badan Pertanahan Daerah untuk menyiapkan lokasi baru bagi warga yang selama ini bertempat tinggal di kaki Gunung Sinabung, mengingat diprediksi bencana erupsi masih akan tetap berlangsung hingga 5 tahun ke depan, " kata Bambang di DPR, Jakarta, Rabu (21 / 2 / 2018). Politikus Partai Golkar itu juga meminta pada Pusat Vulkanologi dan Mitigasi Bencana Geologi, selalu memantau dan memberikan informasi kepada warga tentang situasi dan kondisi status Gunung Sinabung. Ia juga berharap, agar masyarakat dan pengunjung atau wisatawan untuk tidak melakukan aktivitas di dalam radius zona merah. MEDAN – Gunung Sinabung, Kabupaten Karo, Sumatera Utara, kembali erupsi dan meluncurkan awan panas sejauh 2,5 km, Sabtu (27 / 1 / 2018). Selain menyemburkan abu vulkanis setinggi 300 meter, erupsi juga mengakibatkan gempa vulkanis. Pihak PVMBG terus memantau aktivitas gunung Sinabung yang terletak di Kabupaten Karo, Sumatera Utara itu dari CCTV dan Thermal Cam milik PVMBG Badan Geologi. Armen mengimbau masyarakat dan wisatawan jangan melakukan aktivitas di dalam radius 3 km dari puncak, dan dalam jarak 7 km untuk sektor Selatan-Tenggara, di dalam jarak 6 km untuk sektor Tenggara - Timur, serta di dalam jarak 4 km untuk sektor Utara - Timur Sinabung. Medan, Sumatera Utara (ANTARA News) - Gunung Sinabung di Kabupaten Karo meletus pada Senin pukul 08.53 WIB, letusannya lebih besar dari erupsi sebelumnya menurut petugas Pengamat Gunung Sinabung di Pusat Vulkanologi dan Mitigasi Bencana Geologi (PVMBG) M Nurul Saori. Letusan pertama pukul 08.53 WIB cukup kuat sehingga memunculkan semburan awan panas hingga 5.000 meter lebih, lalu disusul beberapa erupsi dengan intensitas lebih rendah menurut Nurul. Semburan awan panas akibat erupsi gunung api tersebut menyebar hingga 4,9 km ke arah selatan dan 3,5 km ke arah timur dan tenggara gunung.

Gambar 4. 7 Hasil Ringkasan Story 27

Sutopo mengimbau agar masyarakat dan pengunjung tidak melakukan aktivitas di dalam radius 3 Km dari puncak, dan dalam jarak 7 Km untuk sektor Selatan - Tenggara, di dalam jarak 6 km untuk sektor Tenggara - Timur, serta di dalam jarak 4 km untuk sektor Utara - Timur G. Sinabung. Suara.com - Gunung Sinabung di Sumatera Utara erupsi untuk kesekian kalinya, pada Senin (19 / 2 / 2018). Menanggapi kejadian itu, Ketua DPR Bambang Soesatyo meminta Badan Nasional Penanggulangan Bencana pusat, berkoodinasi dengan BNPB daerah untuk memperbaiki sarana dan prasarana yang terkait dengan keselamatan dan kesehatan warga di Kaki Gunung Sinabung. " Kita meminta Pemerintah Daerah dan Badan Pertanahan Daerah untuk menyiapkan lokasi baru bagi warga yang selama ini bertempat tinggal di kaki Gunung Sinabung, mengingat diprediksi bencana erupsi masih akan tetap berlangsung hingga 5 tahun ke depan, " kata Bambang di DPR, Jakarta, Rabu (21 / 2 / 2018). MEDAN – Gunung Sinabung, Kabupaten Karo, Sumatera Utara, kembali erupsi dan meluncurkan awan panas sejauh 2,5 km, Sabtu (27 / 1 / 2018). Medan, Sumatera Utara (ANTARA News) - Gunung Sinabung di Kabupaten Karo meletus pada Senin pukul 08.53 WIB, letusannya lebih besar dari erupsi sebelumnya menurut petugas Pengamat Gunung Sinabung di Pusat Vulkanologi dan Mitigasi Bencana Geologi (PVMBG) M Nurul Saori. Letusan pertama pukul 08.53 WIB cukup kuat sehingga memunculkan semburan awan panas hingga 5.000 meter lebih, lalu disusul beberapa erupsi dengan intensitas lebih rendah menurut Nurul.

Gambar 4. 8 Hasil Ringkasan Sistem Story 27

#### 4.3.4 Percobaan *Fold 4*

Tabel 4. 10 menampilkan hasil statistik dokumen dari hasil peringkasan manusia dan peringkasan sistem untuk berita multi dokumen pada story 31-40.

Tabel 4. 10 Statistik Jumlah Kata Dalam Ringkasan *Fold 4*

List Story	kata unik pada ringkasan manusia	kata unik pada ringkasan mesin	kata unik pada ringkasan mesin dan manusia
Story 31	340	300	140
Story 32	280	151	129
Story 33	282	403	282
Story 34	132	165	132
Story 35	155	149	116
Story 36	353	311	267
Story 37	222	128	49
Story 38	312	133	133
Story 39	275	159	144
Story 40	415	159	144

Selanjutnya adalah menghitung nilai ROUGE-1, *Presicion*, dan *F1-Score* pada setiap story dengan persamaan 4.1, 4.2, dan 4.3. Kemudian hasil dari perhitungan ringkasan pada *Fold 2* ditunjukkan pada Tabel 4. 11 dibawah ini

Tabel 4. 11 Hasil Perhitungan ROUGE-1 *Fold 4*

List Story	<i>Recall</i>	<i>Presicion</i>	<i>F-Measure</i>
Story 31	0.411	0.466	0.437
Story 32	0.460	0.854	0.598
Story 33	1.0	0.699	0.823
Story 34	1.0	0.8	0.889
Story 35	0.748	0.778	0.763
Story 36	0.756	0.858	0.804
Story 37	0.220	0.382	0.280
Story 38	0.426	1.0	0.597
Story 39	0.523	0.905	0.663
Story 40	0.662	0.813	0.730

Gambar 4. 9 dan Gambar 4. 10 adalah hasil peringkasan dari story 34 dengan query “sea”, “games”, “timnas”, “indonesia”, “basket” dengan hasil nilai *Recall* 1.0

JUARA.NET - Timnas basket putra Indonesia mengawali perjuangan pada babak penyisihan Grup B SEA Games 2017 dengan kemenangan 89 - 73 atas Vietnam di MABA Stadium, Minggu (20 / 8 / 2017) pagi. Arki Dikania Wisnu menjadi pencetak poin terbanyak pada pertandingan ini dengan torehan 17 poin, diikuti Ebrahim Enguio Lopez, dan Abraham Damar Grahita yang menyumbang 11 dan 10 poin. Mario Wuysang Cs unggul 17 - 15 pada kuartir pertama. Pada kuartir tiga, Indonesia meninggalkan jauh Vietnam dengan skor 75 - 52. Namun, pada kuartir keempat permainan timnas memburuk. JUARA.NET, SURABAYA - CLS Knights Indonesia resmi merekrut pebasket naturalisasi, Ebrahim Enguio Lopez atau biasa disapa Biboy, untuk mengarungi ajang ASEAN Basketball League (ABL) 2017 - 2018. Biboy yang berasal dari Filipina sudah diperkenalkan ke penonton pada laga antara CLS Knights dan Formosa Dreamers di GOR Kertajaya, Surabaya, Sabtu (18 / 11 / 2017).

Gambar 4. 9 Hasil Ringkasan Manusia Story 34

JUARA.NET - Timnas basket putra Indonesia mengawali perjuangan pada babak penyisihan Grup B SEA Games 2017 dengan kemenangan 89 - 73 atas Vietnam di MABA Stadium, Minggu (20 / 8 / 2017) pagi. Arki Dikania Wisnu menjadi pencetak poin terbanyak pada pertandingan ini dengan torehan 17 poin, diikuti Ebrahim Enguio Lopez, dan Abraham Damar Grahita yang menyumbang 11 dan 10 poin. JUARA.NET, SURABAYA - CLS Knights Indonesia resmi merekrut pebasket naturalisasi, Ebrahim Enguio Lopez atau biasa disapa Biboy, untuk mengarungi ajang ASEAN Basketball League (ABL) 2017 - 2018. Biboy yang berasal dari Filipina sudah diperkenalkan ke penonton pada laga antara CLS Knights dan Formosa Dreamers di GOR Kertajaya, Surabaya, Sabtu (18 / 11 / 2017).

Gambar 4. 10 Hasil Ringkasan Sistem Story 34

#### 4.3.5 Percobaan *Fold 5*

Tabel 4. 12 menampilkan hasil statistik dokumen dari hasil peringkasan manusia dan peringkasan sistem untuk berita multi dokumen pada story 41-50.

Tabel 4. 12 Statistik Jumlah Kata Dalam Ringkasan *Fold 5*

List Story	kata unik pada ringkasan manusia	kata unik pada ringkasan mesin	kata unik pada ringkasan mesin dan manusia
Story 41	191	159	125
Story 42	119	161	73
Story 43	186	153	128
Story 44	278	269	119
Story 45	259	230	185
Story 46	156	255	147
Story 47	216	255	213
Story 48	211	198	185
Story 49	320	479	263
Story 50	126	246	96

Selanjutnya adalah menghitung nilai ROUGE-1, *Precision*, dan *F1-Score* pada setiap story dengan persamaan 4.1, 4.2, dan 4.3. Kemudian hasil dari perhitungan ringkasan pada *Fold 2* ditunjukkan pada Tabel 4. 13 dibawah ini

Tabel 4. 13 Hasil Perhitungan ROUGE-1 *Fold 5*

List Story	<i>Recall</i>	<i>Precision</i>	<i>F-Measure</i>
Story 41	0.654	0.786	0.714
Story 42	0.613	0.453	0.521
Story 43	0.688	0.836	0.755
Story 44	0.428	0.442	0.435
Story 45	0.714	0.804	0.756
Story 46	0.942	0.576	0.715
Story 47	0.986	0.835	0.904
Story 48	0.876	0.934	0.658
Story 49	0.821	0.549	0.658
Story 50	0.761	0.390	0.516

Gambar 4. 11 dan Gambar 4. 12 adalah hasil peringkasan dari story 47 dengan query "konser", "jakarta", "celine", "dion" dengan hasil nilai *Recall* 0,986

Jakarta, CNN Indonesia - - Penyanyi kawakan Celine Dion akan menandai kedatangannya pertama kali di Jakarta lewat konser pada pertengahan tahun ini. Pencetak hit lagu tema film Titanic, My Heart Will Go On itu akan melangsungkan konser pada 7 Juli 2018 di Sentul International Convention Center, Indonesia. Pengumuman konser tersebut dikonfirmasi oleh promotor Sound Rhythm melalui akun media sosial resmi pada Jumat (12/1) pagi. "Telah terkonfirmasi, Celine Dion Live 2018 Tour in Jakarta, Sabtu 7 Juli 2018 di Sentul International Convention Center," tulisnya pada unggahan berupa poster konser tersebut. Jakarta, CNN Indonesia - - Keputusan promotor konser Celine Dion di Jakarta yang mematok harga tiket mulai dari Rp 1,5 juta hingga Rp 12,5 juta menuai protes dari netizen. Harga tersebut dianggap terlalu mahal. Pihak promotor Sound Rhythm mengumumkan harga konser penyanyi Kanada itu melalui akun media sosial mereka, beserta dengan desain tempat duduk konser yang akan berlangsung 7 Juli 2018 di Sentul International Convention Center (SICC) tersebut. Jakarta, CNN Indonesia - - Sebagai musisi yang pernah menjalani konser residensi di Las Vegas, Celine Dion menilai tak ada saran yang perlu diberikan kepada Lady Gaga yang akan menjalani proyek residensinya pada akhir 2018 nanti. Pelantun My Heart Will Go On itu menilai Lady Gaga paham berbagai hal yang perlu dilakukan saat konser residensi selama dua tahun ke depan di Las Vegas. Dia tahu apa yang bakal dilakukan," kata Dion.

Gambar 4. 11 Hasil Ringkasan Manusia Story 47

Pencetak hit lagu tema film Titanic, My Heart Will Go On itu akan melangsungkan konser pada 7 Juli 2018 di Sentul International Convention Center, Indonesia. Pengumuman konser tersebut dikonfirmasi oleh promotor Sound Rhythm melalui akun media sosial resmi pada Jumat (12 / 1) pagi. " Telah terkonfirmasi, Celine Dion Live 2018 Tour in Jakarta, Sabtu 7 Juli 2018 di Sentul International Convention Center, " tulisnya pada unggahan berupa poster konser tersebut. Jakarta, CNN Indonesia - - Keputusan promotor konser Celine Dion di Jakarta yang mematok harga tiket mulai dari Rp 1,5 juta hingga Rp 12,5 juta menuai protes dari netizen. Pihak promotor Sound Rhythm mengumumkan harga konser penyanyi Kanada itu melalui akun media sosial mereka, beserta dengan desain tempat duduk konser yang akan berlangsung 7 Juli 2018 di Sentul International Convention Center (SICC) tersebut. Jakarta, CNN Indonesia - - Sebagai musisi yang pernah menjalani konser residensi di Las Vegas, Celine Dion menilai tak ada saran yang perlu diberikan kepada Lady Gaga yang akan menjalani proyek residensinya pada akhir 2018 nanti. Pelantun My Heart Will Go On itu menilai Lady Gaga paham berbagai hal yang perlu dilakukan saat konser residensi selama dua tahun ke depan di Las Vegas.

Gambar 4. 12 Hasil Ringkasan Sistem Story 47

#### 4.4 Pembahasan

Pada Tabel 4. 5 dapat ditunjukkan bahwa nilai *Recall* terbesar pada data *Fold 1* terdapat pada story 7 dan 8 dengan nilai 1. Kemudian pada Tabel 4. 7 nilai *Recall* terbesar pada data *Fold 2* terdapat pada story 11 dan 18, yaitu 1. Kemudian pada Tabel 4. 9 nilai *Recall* terbesar pada *Fold 3* terdapat pada story 27 dan 28 dengan nilai 1. Kemudian pada Tabel 4. 11 nilai *Recall* terbesar pada *Fold 4* terdapat pada story 33 dan 34 dengan nilai 1. Yang terakhir pada Tabel 4. 13 nilai *Recall* terbesar terdapat pada story 47 dengan nilai 0.986. Sedangkan nilai terendah dari *Recall* untuk masing-masing *Fold* adalah *Fold 1* story 2 dengan nilai 0,472, *Fold 2* story 13 dengan nilai 0.747, *Fold 3* story 24 dengan nilai 0.839, *Fold 4* story 37 dengan nilai 0.220, *Fold 5* story 44 dengan nilai 0,428.

Selanjutnya, nilai *Presicion* terbesar pada *Fold 1* terdapat pada story 2 dengan nilai 0.916. Kemudian pada *Fold 2* nilai *Presicion* terbesar terdapat pada story 13 dengan nilai 0.813. Pada *Fold 3* nilai *Presicion* terbesar terdapat pada story 22 dengan nilai 0,632. Pada *Fold 4* nilai *Presicion* terbesar terdapat pada story 38 dengan nilai 1. Kemudian pada *Fold 5* nilai *Presicion* terbesar terdapat pada story

48 dengan nilai 0,934. Sedangkan nilai terendah dari *Precision* untuk masing-masing *Fold* adalah *Fold 1* story 9 dengan nilai 0.284, *Fold 2* story 18 dengan nilai 0.361, *Fold 3* story 28 dengan nilai 0.321, *Fold 4* story 37 dengan nilai 0.382, *Fold 5* story 50 dengan nilai 0.390.

Nilai *F-Measure* terbesar pada *Fold 1* terdapat pada story 6 dengan nilai 0,747. Kemudian nilai *F-Measure* terbesar pada *Fold 2* terdapat pada story 12 dengan nilai 0,797. Selanjutnya nilai *F-Measure* terbesar pada *Fold 3* terdapat pada story 22 sebesar 0,737. Nilai *F-Measure* terbesar pada *Fold 4* terdapat pada story 34 sebesar 0,889. Kemudian nilai *F-Measure* terbesar pada *Fold 5* terdapat pada story 47 dengan nilai sebesar 0.904. Sedangkan nilai terendah dari *F-Measure* untuk masing-masing *Fold* adalah *Fold 1* story 4 dengan nilai 0.409, *Fold 2* story 14 dengan nilai 0.520, *Fold 3* story 28 dengan nilai 0.486, *Fold 4* story 37 dengan nilai 0.280, *Fold 5* story 44 dengan nilai 0.435.

Tabel 4. 14 Nilai Rata-Rata Nilai ROUGE-1 pada 5 *Fold*

<b>List Fold</b>	<b>Recall</b>	<b>Precision</b>	<b>F-Measure</b>
<i>Fold 1</i>	0.797	0.503	0.580
<i>Fold 2</i>	0.870	0.558	0.663
<i>Fold 3</i>	0.946	0.487	0.634
<i>Fold 4</i>	0.620	0.755	0.658
<i>Fold 5</i>	0.748	0.660	0.663

Tabel 4. 14 menunjukkan nilai rata-rata hasil akhir dari sistem yang dibangun dengan 5 *Fold*. Hasil menunjukkan bahwa *Fold 1* memiliki nilai rata-rata *Recall* adalah 0.6797, rata-rata *Precision* sebesar 0.503, dan nilai rata-rata *F-Measure* mencapai 0.580. Selanjutnya, *Fold 2* memiliki nilai rata-rata *Recall* sebesar 0.870, rata-rata *Precision* sebesar 0.558, dan nilai rata-rata *F-Measure* adalah 0.663. *Fold 3* menunjukkan bahwa nilai rata-rata *Recall* adalah 0.948, rata-

rata *Precision* mencapai 0.478, dan nilai rata-rata *F-Measure* adalah 0.634. Pada *Fold 4*, terdapat nilai rata-rata *Recall* sebesar 0.620, rata-rata *Precision* sebesar 0.755, dan nilai rata-rata *F-Measure* mencapai 0.658. Pada *Fold 5*, terdapat nilai rata-rata *Recall* sebesar 0.748, rata-rata *Precision* sebesar 0.660, dan nilai rata-rata *F-Measure* mencapai 0.663.

Hasil tersebut dapat digunakan untuk mewakili seberapa relevan hasil dari ringkasan sistem. Hasil *Recall* dapat digunakan untuk mengukur relevansi hasil ringkasan sistem dengan ringkasan manusia. Semakin tinggi nilai *Recall*, semakin relevan hasil ringkasan sistem. Berdasarkan uji coba sistem, terbukti bahwa sistem dapat melakukan peringkasan otomatis dari berita multi dokumen. Proses ini tidak hanya efisien tetapi juga menghemat waktu dibandingkan dengan peringkasan manual yang memakan waktu lebih lama. Sehingga, kehadiran ringkasan berita ini memberikan kemudahan dan efisiensi dalam memahami informasi dari beberapa dokumen sekaligus.

Hal ini sesuai dengan tujuan sistem peringkasan ini yaitu untuk membantu memperoleh informasi yang penting dan relevan. Memahami peringkasan dapat membantu dalam memahami konteks dan informasi yang ada dalam berita. Terutama jika terdapat banyak berita yang membahas suatu topik, sistem peringkasan multi dokumen ini dapat berfungsi sebagai *filter* agar tidak kehilangan informasi berharga. Sistem ini diharapkan dapat memberikan kontribusi dalam membantu orang lain sebagai bagian dari pelaksanaan perintah Allah subhanahu wa ta'ala yang mendorong kebaikan dan ketakwaan.



Firman Allah S.W.T berfirman pada surat An-Nisa' ayat 101:

وَإِذَا ضَرَبْتُمْ فِي الْأَرْضِ فَلَيْسَ عَلَيْكُمْ جُنَاحٌ أَنْ تَقْصُرُوا مِنَ الصَّلَاةِ إِنَّ خِفْتُمْ أَنْ يُفْتِنَكُمْ الَّذِينَ كَفَرُوا إِنَّ  
الْكَافِرِينَ كَانُوا لَكُمْ عَدُوًّا مُّبِينًا

*“Dan apabila kamu bepergian di muka bumi, maka tidaklah mengapa kamu meng-qashar sembahyang(mu), jika kamu takut diserang orang-orang kafir. Sesungguhnya orang-orang kafir itu adalah musuh yang nyata bagimu.” (QS. An-Nisa' ayat 101)*

Menurut Tafsir Ibnu Katsir meng-*qasar* yaitu meringankan atau meringkas, adakalanya dari segi rakaatnya, misalnya salat yang empat rakaat dijadikan dua rakaat, seperti yang disimpulkan oleh jumur ulama dari ayat ini. Mereka menjadikannya sebagai dalil salat qasar dalam perjalanan, sekalipun mereka masih berselisih pendapat mengenainya. Karena di antara mereka ada yang mengatakan bahwa perjalanan yang dilakukan harus mengandung ketaatan, seperti berjihad, atau haji atau umrah, atau mencari ilmu atau ziarah, atau lain-lainnya.

Sistem peringkasan teks multi dokumen ini adalah upaya untuk mengimplementasikan ayat tersebut tentang konteks meringkas berita. Dalam penelitian ini, meng-*qasar* dapat dianalogikan seseorang dapat melakukan peringkasan karena terdapat permasalahan, yaitu informasi yang berlebihan (*overloaded information*). Informasi yang berlebihan ini dapat membuat pembaca kesulitan memahami inti berita. Dengan demikian, sistem peringkasan yang sudah dibuat diharapkan dapat mengurangi risiko kesalahpahaman terkait informasi berita, karena pembaca dapat dengan mudah memahami inti dari berita melalui ringkasan yang disediakan oleh sistem.

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Hasil uji coba sistem peringkasan teks multi dokumen berbahasa Indonesia dengan metode Sentence Scoring dan SVM pada 50 data multi dokumen menunjukkan bahwa sistem mampu mendapatkan hasil yang relevan dalam mengekstrak story. Hal ini dibuktikan dengan nilai relevansi rata-rata *Recall* mencapai 0.946, *Presicion* rata-rata sebesar 0.487, dan *F-Measure* rata-rata mencapai 0.634. Nilai *Recall* yang tinggi menunjukkan bahwa sistem mampu mengidentifikasi informasi yang relevan dari berbagai dokumen dengan baik. Nilai Presisi yang rendah menunjukkan bahwa sistem masih memiliki kemungkinan untuk memasukkan informasi yang tidak relevan ke dalam ringkasan. Nilai *F-Measure* yang sedang berarti bahwa sistem dapat mengekstrak informasi penting dari berita dengan cukup baik, tetapi masih ada beberapa kekurangan. Nilai *F-Measure* tersebut menunjukkan bahwa sistem masih memiliki kemungkinan untuk memasukkan informasi yang tidak penting ke dalam ringkasan, atau tidak memasukkan informasi penting yang seharusnya ada dalam ringkasan.

Secara keseluruhan, hasil uji coba menunjukkan bahwa sistem peringkasan teks multi dokumen berbahasa Indonesia dengan metode Sentence Scoring dan SVM memiliki performa yang baik dalam mengekstrak informasi yang relevan dari berbagai dokumen. Berdasarkan hasil ini, sistem dapat digunakan untuk menghasilkan ringkasan teks multi dokumen yang informatif dan akurat.

## 5.2 Saran

Berdasarkan hasil penelitian, dapat disimpulkan bahwa sistem peringkasan teks otomatis berbahasa Indonesia dengan metode Sentence Scoring dan SVM memiliki performa yang baik dalam mengekstrak informasi penting dari berita. Namun, sistem ini masih memiliki beberapa kekurangan yaitu, nilai *Recall* yang rendah menunjukkan bahwa sistem masih memiliki kemungkinan untuk memasukkan informasi yang tidak penting ke dalam ringkasan. Kemudian Sistem belum dapat menghasilkan ringkasan yang koheren dan kohesif, terutama untuk berita yang panjang dan kompleks.

Oleh karena itu, Untuk meningkatkan performa sistem, peneliti dapat mempertimbangkan beberapa saran yaitu, Menambahkan dataset yang lebih beragam untuk memperkuat validitas dan keberagaman hasil analisis. Kemudian mempertimbangkan untuk menambahkan fitur pada Sentence Scoring yang terkait dengan hubungan makna antar kalimat dan fitur semantik. Serta Mengembangkan algoritma yang dapat menghasilkan ringkasan yang koheren dan kohesif, terutama untuk berita yang panjang dan kompleks.

## DAFTAR PUSTAKA

- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., D., E., B., J., & Kochut, K. (2017). Text Summarization Techniques: A Brief Survey. *International Journal of Advanced Computer Science and Applications*, 8(10). <https://doi.org/10.14569/ijacsa.2017.081052>
- Anandarajan, M., Hill, C., & Nolan, T. (2018). Text Pre-Processing. In *Advances in Analytics and Data Science* (pp. 45–59). [https://doi.org/10.1007/978-3-030-85085-2\\_3](https://doi.org/10.1007/978-3-030-85085-2_3)
- Aulia, T. M. P., Jamaludin, A., & ... (2021). Extractive Text Summerization Pada Berita Berbahasa Indonesia Menggunakan Algoritma Support Vector Machine. *J-SAKTI (Jurnal Sains ...)*, 5(September), 727–735. <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/371>
- Belwal, R. C., Rai, S., & Gupta, A. (2022). Extractive text summarization using clustering-based topic modeling. *Soft Computing*. <https://doi.org/10.1007/s00500-022-07534-6>
- Christian, H., Agus, M. P., & Suhartono, D. (2016). Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications*, 7(4), 285. <https://doi.org/10.21512/comtech.v7i4.3746>
- Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, 47(1), 1–66. <https://doi.org/10.1007/s10462-016-9475-9>
- Hayatin, N., Marthasari, G. I., & Anggraini, S. (2018). Improvement of cluster importance algorithm with sentence position for news summarization. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2018-October, 483–488. <https://doi.org/10.1109/EECSI.2018.8752633>
- Kurniawan, K., & Louvan, S. (2018). I NDO S UM : A New Benchmark Dataset for Indonesian Text Summarization. *2018 International Conference on Asian Language Processing (IALP)*, 215–220.
- Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2), 159–165. <https://doi.org/10.1147/rd.22.0159>
- Mandal, S., Achary, P., Phalke, S., Poorvaja, K. V. K., & Kulkarni, M. (2021). Extractive Text Summarization Using Supervised Learning and Natural

Language Processing. *2021 International Conference on Intelligent Technologies, CONIT 2021*, 1–7. <https://doi.org/10.1109/CONIT51480.2021.9498322>

Moratanch, N., & Chitrakala, S. (2017). A survey on extractive text summarization. *International Conference on Computer, Communication, and Signal Processing: Special Focus on IoT, ICCOSP 2017*. <https://doi.org/10.1109/ICCCSP.2017.7944061>

Naufal Ammar, A., & Suyanto. (2020). Peringkasan Teks Ekstraktif Menggunakan Binary Firefly Algorithm. *Indonesia Journal of Computing*, 5(2), 31–42. <https://doi.org/10.21108/indojc.2020.5.2.440>

Qaroush, A., Abu Farha, I., Ghanem, W., Washaha, M., & Maali, E. (2021). An efficient single document Arabic text summarization using a combination of statistical and semantic features. *Journal of King Saud University - Computer and Information Sciences*, 33(6), 677–692. <https://doi.org/10.1016/j.jksuci.2019.03.010>

Rahul, Adhikar, S., & Monika. (2020). NLP based Machine Learning Approaches for Text Summarization. *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020, Iccmc*, 535–538. <https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00099>

Rainarli, E., & Dewi, K. E. (2018). Relevance Vector Machine for Summarization. *IOP Conference Series: Materials Science and Engineering*, 407(1). <https://doi.org/10.1088/1757-899X/407/1/012075>

Sabuna, P. M., & Setyohadi, D. B. (2018). Summarizing Indonesian text automatically by using sentence scoring and decision tree. *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017, 2018-Janua*, 1–6. <https://doi.org/10.1109/ICITISEE.2017.8285473>

Saragih, P. S., Witarsyah, D., Hamami, F., & MacHado, J. M. (2021). Sentiment Analysis of Social Media Twitter with Case of Large Scale Social Restriction in Jakarta using Support Vector Machine Algorithm. *2021 International Conference Advancement in Data Science, E-Learning and Information Systems, ICADEIS 2021, 19(January 2020)*, 0–5. <https://doi.org/10.1109/ICADEIS52521.2021.9701961>

Shirwandkar, N. S., & Kulkarni, S. (2018). Extractive Text Summarization Using Deep Learning. *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*, 1–5. <https://doi.org/10.1109/ICCUBEA.2018.8697465>

Somantri, G. P., Komarudin, A., & Ilyas, R. (2018). Peringkasan Teks Otomatis

Berita Berdasarkan Klasifikasi Kalimat Menggunakan Support Vector Machine. *Prosiding SNATIF*, 57–62.

Suyuti, J. al-D. al-, & Mahalli, J. al-D. al-. (2010). *Tafsir Jalalain*. Sinar Baru Algensindo.

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

Verma, P., Pal, S., & Om, H. (2019). *A Comparative Analysis on Hindi and English Extractive*. 18(3).

Widodo, W., Nugraheni, M., & Sari, I. P. (2021). A comparative review of extractive text summarization in Indonesian language. *IOP Conference Series: Materials Science and Engineering*, 1098(3), 032041. <https://doi.org/10.1088/1757-899x/1098/3/032041>

Winters-Hilt, S., Yelundur, A., McChesney, C., & Landry, M. (2006). Support Vector Machine implementations for classification & clustering. *BMC Bioinformatics*, 7(SUPPL.2), 1–18. <https://doi.org/10.1186/1471-2105-7-S2-S4>

Yovellia Londo, G. L., Kartawijaya, D. H., Ivaryani, H. T., Yohanes Sigit, P. W. P., Muhammad Rafi, A. P., & Ariyandi, D. (2019). A Study of Text Classification for Indonesian News Article. *Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT 2019*, 205–208. <https://doi.org/10.1109/ICAIIIT.2019.8834611>

# **LAMPIRAN**

## LAMPIRAN

### Lampiran I Data 50 data multi dokumen

id_story	judul
1	[' Mati Dalam Operasi Antinarkotika Filipina Pro', 'Perang Duterte Lawan Narkoba Dipergencar', 'Filipina Diserang Isu Pelanggaran Ham', 'Rodrigo Duterte Diselidiki Mahkamah Internasional']
2	['Hp Luncurkan Kamera Instan Sprocket In ', 'Jajal Printer Portable Canon Selphy Cp', 'Polaroid Dan Fujifilm Bersengketa Akibat Format Sq']
3	['Arkavidia Itb Akan Fokus Bahas Perkembangan Sta', 'Visio Incubator Umumkan Startup Terpilih Untuk ', 'Clapham Startupfest Segera Hadir Kembali Di M']
4	['Asus Xiu Laptop Multimedia Gaming Ready', 'Asus Rilis Laptop Gim Seharga Motor Sport', 'Asus Ramu Notebook Rog Gvik']
5	['Tesla Akan Tanam Teknologi Otonom Di Semua Mobil L', 'Tiongkok Siapkan Jutaan Dolar Untuk Ganggu Pasar M']
6	['Asia Pacific Predator League Resmi Dimulai', 'Acer Nitro Spin Laptop Gaming In Terbaru', 'Philips Sedang Garap Monitor K']
7	['Angin Meningkatkan Peran Perempuan Di Dunia Teknol', 'Rayakan Agustus Go Jek Sumbang Rp Miliar Ke ']
8	['Juni Tesla Model X Masuk Indonesia', 'Bmw Punya Seri Listrik Untuk Lawan Tesla Model ', 'Tesla Model Mulai Diproduksi']
9	['Asus Xiu Laptop Multimedia Gaming Ready', 'Asus Rilis Laptop Gim Seharga Motor Sport', 'Monitor Gaming Terbaru Asus Andalkan Layar Curved', 'Asus Ramu Notebook Rog Gvik', 'Amd Luncurkan Prosesor Ryzen Untuk Laptop', 'Mouse Gaming Msi Clutch Gm']
10	['Dua Smartphone Zenfone Baru Tawarkan Solusi Bersel', 'Asus Diam Diam Rilis Zenfone Selfie Lite Di Mala', 'Samsung Galaxy A Bidik Generasi Millennial', 'Asus Zenfone Selfie Lite Tiba Di Filipina']
11	['Pencarian Korban Longsor Ponorogo Terkendala Medan', ' Orang Masih Hilang Akibat Longsor Di Ponorogo']
12	['Polisi Temukan Potongan Tangan Di Lokasi Carok', 'Carok Massal Di Madura Tewaskan Tiga Orang', 'Polisi Ringkus Pelaku Carok Massal Di Madura']
13	['Smartphone Flip Premium Samsung W Resmi Dirili', 'Samsung Luncurkan Flip Phone Android W', 'Microsoft Office Versi Android Kini Sudah Tersedia']
14	['Vivo X Plus Ud Akhirnya Diluncurkan', 'Vivo V Plus Tiba Di India', 'Vivo Rilis Teaser Varian X Mars Edition']
15	['Nintendo Targetkan Untuk Memproduksi Dua Juta Unit', 'Nintendo Switch Konsol Game Paling Ditunggu Hadir ', 'Nintendo Switch Sudah Terjual Juta Unit', 'Nintendo Akan Kembali Produksi Nes Classic']



	Edition', 'Beli Game Nintendo Switch Di Eshop Dengan Paypal', 'Nintendo Umumkan Jadwal Rilis Game Baru Untuk Swit', 'Penjualan Nintendo Switch Sukses Melampaui Pencapa']
16	['Seagate Keluarkan Harddisk Kapasitas Tb', 'Corsair Gelar Kompetisi Modifikasi Case Komputer T', 'Menjajal Hard Disk Eksternal Lacie Dji Copilot']
17	['Google Assistant Kini Bisa Kenali Lagu Seperti Sha', 'Deezer Luncurkan Songcatcher', 'Apple Resmi Akuisisi Aplikasi Penebak Judul Lagu S']
18	['Jokowi Siapkan Regulasi Mobil Listrik', 'Presiden Jokowi Mengaku Deg Degan Jajal Mobil List']
19	['Jokowi Hormati Putusan Ma Soal Perda', 'Jokowi Resmikan Tol Becakayu Proyek Mangkrak Dua D', 'Resmikan Tol Jokowi Bertolak Ke Ujung Selatan Pula', 'Presiden Jokowi Resmikan Tol Pertama Di Sumatera S', 'Kredit Ultra Mikro Untuk Ratakan Kesejahteraan', 'Pembangunan Infrastruktur Jadi Kunci Indonesia Men']
20	['Militer Korsel Ancam Balas Dendam Serangan Korut', 'Di Tengah Ancaman Korut Jepang As Latihan Maritim ']
21	['Tujuh Suku Papua Ingin Dialog Dengan Freeport', 'Jokowi Untung Kita Punya Menteri Susi Yang Tenggel', 'Mike Pence Kunjungi Indonesia April']
22	['Pemerintah Naikkan Cukai Rokok Persen', 'Ylki Kenaikan Harga Rokok Terlalu Kecil Harusnya J', 'Sri Mulyani Resmikan Harga Rokok Naik Mulai Janu', 'Cukai Resmi Naik Harga Jual Eceran Rokok Bakal Mel']
23	['Pertamina Realisasikan Wilayah Bbm Satu Harga', 'Iri Bbm Satu Harga Pgn Ingin Harga Gas Sama Rata', 'Esdm Bbm Satu Harga Bukan Masalah Untung Rugi Tapi', 'Salurkan Bbm Satu Harga Pertamina Butuh Rp T', ' Bbm Satu Harga Masuk Ke Daerah Terpencil', 'Dpr Dukung Kebijakan Energi Untuk Rakyat']
24	['Kisruh Impor Beras Dpr Minta Pemerintah Validasi D', 'Bps Aktivitas Ekspor Impor Beras Melambung Hingga ', 'Pengumuman Impor Beras Seret Harga Beras Turun Rp', 'Enggar Impor Beras Ton Tanpa Apbn', 'Sekjen Partai Komunis Vietnam Bahas Zona Perdagangan']
25	['Gara Gara Pemilu As Harga Emas Dunia Turun', 'Akhir Pekan Rupiah Dan Harga Komoditas Warnai Gera', 'Euro Melemah Usai Catalonia Ingin Pisah Dari Spanyol', 'Hari Ini Harga Emas Turun Rp Menjadi Rp ', 'Emas Tertekan Penguatan Saham Dan Dolar As']
26	['Anies Enggan Beberkan Rencana Atasi Banjir Jakarta', 'Banjir Karena Penyempitan Kali Anies Minta Warga G', 'Antisipasi Banjir Kiriman Pemprov Dki Siagakan Sel', 'Pemprov Dki Berencana Kaji Ulang Intensif Pasukan ', 'Demi Selamatkan Jakarta Dari Banjir Anies Rela Lak']
27	['Gunung Sinabung Kembali Erupsi', 'Ketua Dpr Minta Sarana Di Kawasan Sinabung Mulai D', 'Gunung Sinabung Semburkan Awan Panas', 'Gunung Sinabung Meletus']

28	['Dirut Pt Garam Terancam Tahun Penjara', 'Polri Tangkap Dirut Pt Garam']
29	['Sri Mulyani Yakin Ekonomi Kuartal Iv Tumbuh Tembus', 'Sri Mulyani Kestabilan Inflasi Adalah Kunci', 'Indonesia Masih Rendah Ketimbang Negara Lain', 'Sri Mulyani Ekonomi Tumbuh Persen Tahun Ini']
30	['Polri Luncurkan Aplikasi Peringatan Dini Gunung Ag', 'Gunung Agung Bali Meletus', 'Gunung Agung Meletus Lagi', 'Gubernur Pastika Ajak Warga Tidak Panik Hadapi Sta', 'Banyuwangi Belum Terdampak Abu Vulkanis Gunung Agu']
31	['Aji Santoso Selamat Untuk Persija', 'Liga Mitra Kukar Arema Fc', 'Liga Madura United Arema Fc', 'Teco Kami Menang Berkat Jakmania']
32	['Tanpa Iwan Setiawan Persebaya Raih Tiga Poin Perda', 'Tundukkan Persebaya Kalteng Putra Lolos Ke Babak D']
33	['Persib Pastikan Jamu Arema Di Stadion Gbla', 'Djanur Genjot Persiapan Pasukannya', 'Ridwan Kamil Prediksi Persib Kalahkan Arema Fc ', 'Supardi Akan Doakan Persib Di Tanah Suci']
34	['Tim Basket Putra Indonesia Raih Kemenangan Pada Se', 'Cls Knights Rekrut Biboy Untuk Arungi Abl']
35	['Kartu Merah Melayang Saat Indonesia Kalahkan Filip', 'Tim Putri Voli Indonesia Lolos Ke Final']
36	['Klasemen Perolehan Medali Sea Games Hingga Selasa ', 'Jangan Owi Butet Terus Dong', 'Tontowililiyana Dapat Bonus Rumah Mewah']
37	['Batshuayi Bawa Chelsea Juara Premier League', 'Wenger Bahas Masa Depan Di Arsenal Pada Selasa']
38	['Greysiaapriani Ditarget All England Dan Asian Game', 'Soerjadi Tokoh Bulu Tangkis Indonesia Wafat']
39	['Menpora Medali Paralympic Brasil Kado Haornas ', 'Rp M Untuk Atlet Difabel Peraih Medali Paralimpia']
40	['Demi Ambisi Piala Dunia China U Berkompetisi Di', 'Prediksi Kamerun Vs Chile Juni ', 'Piala Konfederasi Rusia Lewati Hadangan Perdana']
41	['Portugal Sebut Juara Euro Bukan Keberuntungan', 'Kiko Insa Sebut Manajer Umum Arema Fc Pembohong', 'Wenger Yakin Prancis Akan Jadi Kekuatan Baru Sepak']
42	['Indonesia Buka Peluang Jadi Tuan Rumah Olimpiade', 'Jokowi Ingin Indonesia Punya Banyak Kompetisi Olah']
43	['Dani Alves Jadi Kapten Timnas Brasil', 'Tontowililiyana Perpanjang Asa Akhiri Kutukan Pada']
44	['Lima Hal Yang Harus Diketahui Sebelum Menonton Tho', 'Spider Man Kalahkan Empat Tim Avengers', 'Chris Hemsworth Nyaris Menolak Peran Thor']
45	['Taylor Swift Membuat Aplikasi Jejaring Sosial', 'Curhat Emosional Taylor Swift Di Pembukaan Reputat']

46	['Agnéz Mo Siap Tampilkan Lagu Terbaru', 'Lagu Duet Afgan Raisa Raih Triple Platinum', 'Afgan Gelar Konser Di Kuala Lumpur']
47	['Celine Dion Akan Gelar Konser Pertama Di Indonesia', 'Netizen Protes Tiket Konser Dp Mobil Celine Dion', 'Celine Dion Ogah Beri Saran Lady Gaga Soal Konser ']
48	['Tiket Ed Sheeran Ludes Terjual Dalam Sehari', 'Penjual Tiket Palsu Konser Ed Sheeran Di Singapura']
49	['Musisi Pengisi Synchronize Festival Siap Beri', 'Musik Dangdut Masuk Dalam Synchronize Fest ', 'Jimi The Upstairs Berharap Tampil Di Synchronize F', 'Pujian Jokowi Untuk Musisi Yang Tampil Di Synchron']
50	['Raisa Dan Ashanti Ikut Kirim Doa Ke Kahiyang Bobby', 'Dorce Gamalama Beri Kado Lagu Untuk Iriana Dan Kah']

## Lampiran II Rekap hasil peringkasan 50 Story

List Story	kata unik manusia	kata unik mesin	Kata unik mesin dan manusia	<i>Recall</i>	<i>Presicion</i>	<i>F-Measure</i>
Story 1	491	683	411	0,83706	0,60175	0,7001
Story 2	397	336	240	0,60453	0,71428	0,65484
Story 3	287	532	183	0,65827	0,34398	0,45185
Story 4	121	465	100	0,82644	0,21505	0,34129
Story 5	118	316	98	0,83050	0,31012	0,45161
Story 6	330	645	297	0,90000	0,46046	0,60923
Story 7	222	465	222	01.00	0,47741	0,64628
Story 8	177	531	177	01.00	0,33333	00.05
Story 9	434	1407	400	0,92165	0,28429	0,43454
Story 10	362	780	322	0,88950	0,41282	0,56392
Story 11	146	251	146	01.00	0,58167	0,73551
Story 12	401	459	343	0,59375	0,51875	0,55347
Story 13	321	295	240	0,51875	0,56458	0,54097
Story 14	212	414	163	0,53333	0,27292	0,36111
Story 15	518	1165	457	0,6125	0,27222	0,37708
Story 16	215	385	197	0,63611	0,35486	0,45556
Story 17	172	335	146	0,58889	0,30208	0,39931
Story 18	117	324	117	01.00	0,25069	0,36806
Story 19	713	933	628	0,61111	0,46736	0,52986
Story 20	247	298	201	0,56458	0,46806	0,51181
Story 21	219	424	204	0,64653	0,33403	0,44028
Story 22	495	693	438	0,61389	0,43889	0,51181
Story 23	611	1155	610	0,69306	0,36667	0,47917
Story 24	517	702	434	0,58264	0,42917	0,49444
Story 25	417	670	364	0,60556	0,37708	0,46458
Story 26	345	854	344	0,69236	0,27917	0,39792
Story 27	270	540	270	01.00	00.05	0,4625
Story 28	139	432	139	01.00	0,22292	0,3375
Story 29	339	967	337	0,69028	0,24167	0,35833
Story 30	429	807	409	0,66181	0,35139	0,45903
Story 31	340	300	140	0,28542	0,32361	0,30347
Story 32	280	151	129	0,31944	0,59306	0,41528
Story 33	282	403	282	01.00	0,48542	0,57153

Story 34	132	165	132	01.00	00.08	0,61736
Story 35	155	149	116	0,51944	0,54028	0,52986
Story 36	353	311	267	0,525	0,59583	0,55833
Story 37	222	128	49	0,15278	0,26528	0,19444
Story 38	312	133	133	0,29583	01.00	0,41458
Story 39	275	159	144	0,36319	0,62847	0,46042
Story 40	415	159	144	0,45972	0,56458	0,50694
Story 41	191	159	125	0,45417	0,54583	0,49583
Story 42	119	161	73	0,42569	0,31458	0,36181
Story 43	186	153	128	0,47778	0,58056	0,52431
Story 44	278	269	119	0,29722	0,30694	0,30208
Story 45	259	230	185	0,49583	0,55833	0,525
Story 46	156	255	147	0,65417	0,4	0,49653
Story 47	216	255	213	0,68472	0,57986	0,62778
Story 48	211	198	185	0,60833	0,64861	0,45694
Story 49	320	479	263	0,57014	0,38125	0,45694
Story 50	126	246	96	0,52847	0,27083	0,35833