

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada perkembangan teknologi, data teks dengan format digital semakin meningkat, dan diperlukan manajemen untuk menyediakan data tersebut. Bidang ilmu dalam teknologi komputer dapat digunakan salah satu fungsinya untuk mengelola data teks yaitu *text mining*. *Text mining* merupakan proses menganalisis data teks yang tidak terstruktur guna mengidentifikasi pola, kata kunci, serta informasi penting lainnya dalam data. dengan bantuan perangkat lunak. Salah satu pemanfaatan teks mining ini digunakan untuk meringkas teks pada sebuah dokumen, sehingga peringkasan dokumen dapat dilakukan secara manual dan juga secara digital [1].

Peringkasan teks adalah cara untuk membuat ringkasan teks yang menggunakan komputer untuk menyusun ringkasan dari suatu teks. Metode ini berfungsi untuk menyederhanakan dokumen secara otomatis sambil tetap mempertahankan inti informasi yang terdapat dalam teks aslinya. Proses ini melibatkan pemadatan teks dengan menghilangkan informasi yang tidak diperlukan dan mempertahankan bagian yang paling penting agar tetap mewakili keseluruhan dokumen. Hasil dari peringkasan teks berupa rangkuman yang memuat kalimat-kalimat utama serta informasi relevan dari dokumen aslinya. Peringkasan teks dapat menjadi salah satu alternatif untuk mempermudah pembaca memperoleh sebuah informasi dengan waktu yang cepat terutama untuk digunakan dalam *resume* buku pelajaran dalam dunia pendidikan [2].

Pada dunia pendidikan, buku pelajaran merupakan sumber utama yang digunakan oleh siswa untuk memahami materi pembelajaran. Namun, banyaknya informasi yang terkandung dalam buku teks sering kali membuat siswa kesulitan dalam menyerap informasi secara efektif. Hal ini terutama terjadi pada mata pelajaran seperti Sejarah, yang memiliki konten naratif dan deskriptif. Oleh sebab itu, diperlukan metode untuk dapat membantu menyederhanakan informasi agar lebih mudah dipahami tanpa mengurangi esensi penting dari materi yang disampaikan. Berbagai penelitian menunjukkan bahwa literasi memiliki dampak

besar, baik bagi kehidupan individu maupun masyarakat secara keseluruhan. Namun, hingga saat ini, pengembangan literasi dalam sistem pendidikan di Indonesia masih belum berjalan secara optimal [3].

Resume buku sejarah sendiri memiliki manfaat untuk mengefisiensi waktu pada saat mencari informasi saat membaca buku, dengan adanya *resume* dengan sistem ini dapat mempermudah untuk belajar karena tidak perlu membaca teks yang panjang untuk menemukan sebuah informasi.

Salah satu cara untuk mewujudkan kemudahan dengan *resume* ini adalah, dengan mengubah kelemahan dari metode manual di atas menunjukkan perlunya solusi yang lebih efisien. Salah satu cara untuk mengatasinya adalah dengan mengembangkan alat peringkasan teks yang memungkinkan kita memperoleh inti dari suatu informasi tanpa harus membaca keseluruhan teks. Automatic Text Summarization menjadi solusi yang tepat dalam merangkum dokumen serta tetap menjaga kualitas ringkasan. Dengan adanya teknologi ini, efisiensi dalam menyerap informasi, seperti yang terdapat dalam buku pelajaran, dapat meningkat secara signifikan. [4].

Namun demikian, tidak semua metode peringkasan otomatis memberikan hasil yang optimal. Penelitian oleh Zhang et al. (2023) menunjukkan bahwa sekitar 30% hasil ringkasan dari metode ekstraktif populer seperti LexRank dan TextRank mengandung masalah seperti koreferensi yang keliru atau struktur diskursus yang menyesatkan, sehingga makna informasi dapat terganggu. Hal ini menunjukkan bahwa meskipun metode tersebut mempertahankan kalimat asli, hasil ringkasannya belum tentu benar secara semantik maupun koheren, terutama pada teks panjang seperti buku pelajaran [13].

Salah satu metode yang dapat digunakan untuk melakukan peringkasan otomatis adalah *Latent Semantic Analysis* (LSA). Metode ini menggabungkan konsep aljabar dan statistik untuk mengungkap struktur tersembunyi dalam kata, hubungan antar kata, makna, maupun dokumen. *Latent Semantic Analysis* (LSA) telah diterapkan dalam berbagai penelitian, salah satunya adalah peringkasan teks berita berbahasa Indonesia yang dilakukan oleh Saputra [5]. Selain itu, penelitian yang dilakukan oleh Xiong dan Luo [5] membandingkan efektivitas metode LSA

dengan MMR dalam peringkasan teks, di mana hasilnya menunjukkan bahwa LSA memberikan hasil yang lebih baik.

Berdasarkan temuan tersebut, LSA dapat dianggap sebagai metode yang relevan untuk peringkasan teks, khususnya karena kemampuannya dalam menganalisis keterkaitan semantik antar kata dan kalimat. LSA bekerja dengan pendekatan matematis melalui proses *Singular Value Decomposition* (SVD) untuk mereduksi dimensi dan menemukan makna laten dari dokumen. Keunggulan ini membuat LSA cocok digunakan pada teks yang bersifat panjang dan kompleks, seperti buku pelajaran.

Namun demikian, penerapan LSA dalam konteks pendidikan, khususnya pada buku pelajaran, masih jarang ditemukan. Penelitian ini bertujuan untuk mengembangkan sistem peringkasan berbasis LSA untuk buku Sejarah SMA Kelas XI. Sistem ini akan dievaluasi menggunakan precision, recall, akurasi dan divalidasi secara manual oleh pakar. Dengan pendekatan ini, diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan sistem pendukung pembelajaran yang mampu menyederhanakan materi secara efektif tanpa mengurangi esensi isi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka perumusan masalah dalam penelitian ini adalah:

1.2.1 Permasalahan

Buku pelajaran, khususnya pada mata pelajaran Sejarah SMA, sering kali mengandung teks yang panjang dan kompleks. Hal ini dapat menjadi tantangan bagi siswa yang harus menyaring dan memahami informasi yang relevan dari keseluruhan teks. Siswa sering kesulitan dalam membedakan antara informasi yang penting dan yang kurang penting, sehingga mereka membutuhkan waktu lebih lama untuk mempelajari materi. Di sisi lain, guru juga dihadapkan pada tantangan dalam menyampaikan materi secara ringkas dan efisien tanpa mengorbankan esensi pembelajaran yang harus diterima siswa. Penggunaan metode membaca manual dalam belajar untuk mencari informasi pada buku sering kali memakan waktu dan tidak praktis, terutama untuk dokumen dengan jumlah

halaman yang banyak, seperti buku pelajaran. Oleh karena itu, diperlukan solusi yang mampu menghasilkan ringkasan secara otomatis dengan tingkat keakuratan yang tinggi. Penerapan metode *Latent Semantic Analysis* (LSA) menawarkan solusi dalam mengatasi permasalahan ini, karena mampu menganalisis hubungan semantik antar kata di dalam teks dan menyusun ringkasan yang tetap mempertahankan inti dari informasi.[1]

1.2.2 Solusi Permasalahan

Merancang sistem yang memanfaatkan metode *Latent Semantic Analysis* (LSA) untuk menghasilkan ringkasan yang akurat.

1.3 Pertanyaan Penelitian

Berapa nilai akurasi untuk metode LSA dalam menghasilkan ringkasan otomatis dari buku pelajaran sejarah?

1.4 Batasan Masalah

Penelitian ini memiliki beberapa batasan masalah sebagai berikut:

1. Penelitian akan berfokus pada penggunaan metode LSA dalam meringkas buku Pelajaran.
2. Proses meringkas hanya akan mencakup teks tidak termasuk gambar.
3. Ringkasan yang dihasilkan oleh LSA hanya memberikan gambaran umum dari teks, sehingga mungkin terdapat beberapa informasi penting yang terlewatkan.
4. Penelitian ini menggunakan data dari bab 4 buku Sejarah SMA Kelas XI Edisi cetakan pertama, 2021. Dengan data yang berjumlah 137 kalimat dan terdiri dari 2152 kata.

1.5 Tujuan dan Manfaat

1.5.1 Tujuan

Penelitian ini bertujuan untuk:

1. Menerapkan metode *Latent Semantic Analysis* (LSA) sebagai pembuatan ringkasan pada buku Sejarah SMA Kelas XI.

2. Mengembangkan alat bantu belajar berbasis teknologi yang fleksibel yang bisa digunakan oleh siswa dan guru untuk membantu pembelajaran digital dalam mata Pelajaran sejarah

1.5.2 Manfaat

Penelitian ini diharapkan memberikan beberapa manfaat sebagai berikut:

1. Mempermudah dalam memahami materi pelajaran dengan lebih cepat melalui ringkasan yang fokus pada informasi penting.
2. Menghemat waktu dalam mempelajari buku Sejarah, sehingga siswa dapat lebih efisien dalam persiapan ujian atau tugas.
3. Membantu guru dalam menyiapkan materi ajar secara ringkas tanpa mengurangi esensi dari materi yang disampaikan.
4. Menyediakan alat bantu yang dapat digunakan untuk memfasilitasi proses belajar mengajar, terutama saat harus menyederhanakan materi yang panjang dan kompleks.
5. Mengurangi hambatan dalam memahami teks Sejarah yang kompleks terutama untuk siswa dengan tingkat membaca lebih lambat atau kesulitan dalam menyerap informasi.

BAB II

KAJIAN PUSTAKA

2.1 Landasan Teori

2.1.1 Sejarah

Dalam kajian makna kata, "sejarah" berasal dari bahasa Arab *Syajaratun*, yang berarti "pohon kayu". Pohon ini dianalogikan sebagai sesuatu yang tumbuh dari akar, bercabang, serta menghasilkan dahan, daun, bunga, dan buah. Dalam konteks ini, sejarah menggambarkan rangkaian peristiwa yang terjadi secara berkesinambungan dari masa lalu hingga membentuk jejak serta konsekuensi yang terus berkembang. Sementara itu, dalam bahasa Inggris, sejarah disebut *history*, yang berasal dari bahasa Yunani *histori*, yang berarti "pengetahuan yang diperoleh melalui penyelidikan," merujuk pada pemahaman mengenai berbagai peristiwa yang telah terjadi. [5].

2.1.2 Peringkasan teks otomatis

Peringkasan teks secara otomatis (*automatic text summarization*) adalah cara untuk pembuatan ringkasan dari sebuah teks dengan memanfaatkan komputer. Peringkasan teks otomatis juga merupakan proses mengurangi dokumen teks dengan komputer untuk menciptakan sebuah ringkasan yang mempertahankan poin penting dari dokumen asli. Peringkasan teks merupakan salah satu solusi menemukan inti dari teks dokumen ke dalam bentuk deskripsi singkat. [2]

2.1.3 Preprocessing text

Preprocessing text adalah langkah awal dalam mempersiapkan teks agar dapat diolah sebagai data. Sebuah teks tidak bisa langsung digunakan oleh algoritma pencarian, sehingga diperlukan proses *preprocessing text* untuk mengonversi teks menjadi data numerik [2]. Tahapan ini mencakup berbagai proses pembersihan dokumen, seperti berikut:

1. Pemecahan kalimat adalah proses pemecahan teks dokumen menjadi kalimat [6].
2. *Case Folding* adalah proses mengubah semua huruf dalam sebuah kalimat menjadi huruf kecil (*lowercase*). [7].

3. *Tokenizing* adalah proses memecah suatu kalimat menjadi kata-kata tunggal dengan memisahkannya berdasarkan spasi [7].
4. *Stopword Removal* adalah proses menghapus kata-kata yang tidak memiliki makna signifikan atau tidak berkontribusi terhadap isi utama dokumen, sehingga tidak terlalu penting dalam penyusunan sebuah kalimat [7].
5. *Stemming* sendiri adalah proses untuk menghilangkan kata imbuhan pada suatu kata untuk menjadikan kalimat tersebut menjadi menjadi kata dasar [7].

2.1.4 TF-IDF

Setelah tahap *preprocessing* selesai maka proses selanjutnya ialah proses TF-IDF dimana ini merupakan proses awal pembentukan matriks SVD. Proses TF-IDF diawali dengan mengambil semua term. Kemudian dari setiap file akan diambil kata yang dianggap penting kemudian ditambahkan pada list term. Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan teknik menentukan bobot kata (*term*) sebuah dokumen berdasarkan hubungan antar dokumen tersebut. TF-IDF memberikan bobot tinggi pada kata yang sering muncul pada suatu dokumen.[7]

Rumus TF-IDF:

$$TF = \frac{\text{Jumlah kemunculan kata dalam dokumen}}{\text{Total kata dalam dokumen}} \dots\dots\dots (2.1)$$

$$IDF = \log\left(\frac{\text{Total dokumen}}{\text{Dokumen yang mengandung kata}}\right) \dots\dots\dots (2.2)$$

$$TF-IDF = TF \times IDF \dots\dots\dots (2.3)$$

Tf : (Frekuensi kemunculan kata pada setiap dokumen)

D : (Jumlah dokumen)

Df : (Jumlah dokumen yang mengandung sebuah term)

Idf : (Hasil log dari jumlah dokumen keseluruhan dibagi dengan jumlah dokumen yang mengandung sebuah term)

2.1.5 Latent Semantic Analysis (LSA)

Metode Latent Semantic Analysis (LSA) adalah kombinasi dari metode aljabar dan statistik. Teknik ini memunculkan struktur kata yang tersembunyi, antara kata, perasaan atau dokumen. Hasil penelitian tersebut menunjukkan bahwa metode LSA lebih baik. Kelebihan peringkasan dokumen dengan metode LSA menggunakan Singular Value Decomposition (SVD) menurut Ozsoy adalah untuk

menemukan kata dan kalimat yang mirip secara semantik. SVD memiliki kemampuan pengurangan noise, yang membantu untuk meningkatkan akurasi. [5].

2.1.6 Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) Merupakan teknik aljabar linear yang memiliki fungsi dalam pengolahan dokumen maupun teks. SVD dikenal sebagai teknik yang sangat baik untuk penyelesaian masalah persamaan atau matriks, baik singular maupun secara numerik yang mendekati singular. Keunggulan pada SVD adalah kemampuan untuk digunakan pada semua matriks riil berukuran berikut adalah rumus SVD. [7]

Rumus SVD:

$$A = USV^T \dots\dots\dots (2.4)$$

A : (Input dari matrix (m x n))

U : (Nilai dari Words x concept (m x n))

Sk : (Nilai dari Singular, matriks diagonal (n x n))

V T : (Nilai Sentences x concept (n x n))

2.1.7 Confusion Matrix

merupakan metode evaluasi yang digunakan untuk menilai sistem peringkasan otomatis dengan membandingkan ringkasan yang dihasilkan oleh sistem dengan ringkasan yang dibuat oleh chat GPT. Nilai confusion dituliskan dalam bentuk *precision*, *recall*, dan akurasi.

$$Precision = \frac{tp}{tp+fp} \dots\dots\dots (2.5)$$

Precision adalah berapa banyak dokumen yang berhasil diambil oleh sistem.

Rumus perhitungan precision seperti pada Persamaan (2)

$$Recall = \frac{tp}{tp+fn} \dots\dots\dots (2.6)$$

Recall ialah kemampuan untuk mengambil peringkat teratas yang sebagian besar relevan (benar). (2)

$$Akurasi = \frac{tp+tn}{tp+fn+fp+tn} \dots\dots\dots (2.7)$$

Akurasi merupakan salah satu perhitungan evaluasi yang mengkombinasikan antara kalimat yang terpilih oleh manusia, kalimat yang tidak terpilih oleh manusia, dan kalimat hasil pengambil sistem.(2)

2.2 Penelitian Terdahulu

Pada tabel penelitian terdahulu ini menyajikan ringkasan beberapa penelitian terdahulu yang relevan dengan topik ringkasan otomatis menggunakan pendekatan *Latent Semantic Analysis* (LSA).

Tabel 2.1 Penelitian terdahulu

Penulis dan Tahun Publish	Judul Penelitian	Metode atau Pendekatan	Objek Penelitian	Hasil Penelitian
Yohannes Veli Sabatino (2021) [7]	Penerapan <i>Latent Semantic Analysis</i> untuk Pencarian Kemiripan Berita Berdasarkan Relevansi Berita Berdasarkan Kata Kunci <i>Website</i>	LSA dengan <i>threshold</i> pada pencarian relevansi	Berita <i>Website</i>	Akurasi 60.25%, <i>Precision</i> 92.75%, <i>Recall</i> 62.75% berdasarkan <i>confusion matrix</i> .
Kamal Fauzan Navaro (2021) [8]	Deteksi Plagiarisme Artikel Jurnal Menggunakan <i>Latent Semantic Analysis</i> (LSA)	LSA untuk analisis kesamaan dokumen	Artikel Jurnal	Akurasi rata-rata 87.88% dengan 200 data <i>training</i> dan 20 data <i>testing</i> .

Imam Fahrur Rozi et al. (2021) [7]	Otomatisasi Peringkasan Teks Pada Dokumen Hukum Menggunakan Metode <i>Latent Semantic Analysis</i>	LSA dengan SVD untuk peringkasan teks otomatis	Dokumen hukum	<i>Compression rate: 75%</i> (<i>Precision: 53%, Recall: 27%, F-measure: 35%, Accuracy: 71%</i>), 50% (<i>Precision: 54%, Recall: 56%, F-measure: 55%, Accuracy: 75%</i>), 25% (<i>Precision: 51%, Recall: 79%, F-measure: 61%, Accuracy: 75%</i>).
Dewi Fatmarani Surianto et al.(2022) [9]	Implementasi Metode <i>Latent Semantic Analysis</i> Pada Peringkasan Artikel Bahasa Indonesia Menggunakan Pendekatan	LSA dengan teknik Steinberger dan Jezek untuk pemilihan kalimat	Artikel berita Bahasa Indonesia	<i>Precision</i> tertinggi 20% (<i>Compression rate 10%</i>), rata-rata <i>Recall</i> tertinggi 42% (<i>Compression rate 50%</i>), f-

	Steinberger Jezek			<i>measure</i> tertinggi 22% (<i>Compression</i> <i>rate</i> 50%).
Herlina Jayadianti et al.(2020) [10]	<i>Latent</i> <i>Semantic</i> <i>Analysis</i> (LSA) dan Automatic Text Summarization (ATS) Dalam Optimasi Pencarian Artikel Covid- 19	LSA untuk peringkasan teks otomatis	Artikel Covid-19 dari situs berita	Dataset: 200 dokumen latih, 20 dokumen uji; <i>f-measure</i> rata-rata 90.68%, <i>Recall</i> rata- rata 85%, <i>Compression</i> <i>rate</i> 30%; rasio data latih:uji 90:10.

Yohannes Veli Sabatino dalam penelitiannya yang berjudul Penerapan *Latent Semantic Analysis* Untuk Pencarian Kemiripan Berita Berdasarkan Relevansi Berita Berdasarkan Kata Kunci *website* menyatakan Hasil uji coba kesesuaian logika perhitungan dengan *confusion matrix* juga dilakukan, dengan membandingkan hasil perhitungan yang telah dikodekan pada sistem berdasarkan peenerapan LSA yang telah berjalan pada *website* dengan perbandingan 4 macam threshold menghasilkan bahwa *website* memiliki *accuracy* sebesar 60,25%, *precision* sebesar 92,75% dan *recall* sebesar 62,75% [9]

Navaro *et al.*, (2023) dalam penelitian yang berjudul Deteksi Plagiarisme Artikel Jurnal menggunakan *Latent Semantic Analysis* (LSA) menyatakan bahwa Aplikasi deteksi plagiarisme ini dibangun dengan mengimplementasikan metode *Latent Semantic Analysis* (LSA). Berdasarkan uraian diatas maka dibuatlah Analisa Kesamaan Dokumen Untuk Deteksi Plagiarisme Jurnal Menggunakan LSA Berbasis Web. Pengujian sistem dilakukan untuk mengukur kinerja dari metode

Latent Semantic Analysis (LSA) yang digunakan pada sistem pada 200 data *training* dan 20 data *testing* maka didapatkan rata – rata persentase akurasi sebesar 87.88%. Persentase akurasi yang didapatkan cukup besar sehingga dapat dikatakan bahwa sistem yang dibuat cukup bagus [10].

Imam Fahrur Rozi et al., pada penelitian yang berjudul Otomatisasi Peringkasan Teks Pada Dokumen Hukum Menggunakan Metode *Latent Semantic Analysis*, disebutkan bahwa hasil pengujian terhadap 10 dokumen yang diringkaskan oleh pakar menunjukkan nilai *precision*, *recall*, *f-measure*, dan *accuracy* secara berurutan untuk peringkasan teks otomatis menggunakan metode *Latent Semantic Analysis*. Pada *compression rate* 75%, hasil yang diperoleh adalah 53%, 27%, 35%, dan 71%. Sedangkan untuk *compression rate* 50%, nilai yang didapat adalah 54%, 56%, 55%, dan 75%. Sementara itu, pada *compression rate* 25%, hasil yang dicapai adalah 51%, 79%, 61%, dan 75%. Berdasarkan hasil penelitian dan pengujian tersebut, dapat disimpulkan bahwa metode *Latent Semantic Analysis* dapat digunakan dalam peringkasan dokumen hukum. [7]

Dewi Fatmarani Surianto et al., menyatakan bahwa setelah melakukan eksperimen dalam peringkasan teks artikel berita berbahasa Indonesia dengan menggunakan metode *Latent Semantic Analysis* (LSA) serta teknik *steinberger* dan *jezek* untuk pemilihan kalimat, diperoleh hasil yang bervariasi berdasarkan perbedaan nilai *compression rate* yang diterapkan. *precision* tertinggi, yaitu 20%, dicapai saat *compression rate* berada di angka 10%. Sementara itu, rata-rata *recall* tertinggi sebesar 42% ditemukan pada *compression rate* 50%. Hal yang sama juga terlihat pada nilai *f-measure*, di mana nilai tertinggi yang diperoleh mencapai 22% ketika ringkasan menggunakan *compression rate* sebesar 50%. [11]

Herlina Jayadianti et al., dalam penelitiannya menyatakan bahwa sistem yang dikembangkan dapat meningkatkan efisiensi pencarian artikel terkait Covid-19 dari berbagai situs berita, seperti Kompas.com, Liputan6.com, Klikdokter.com, dan Detikhealth.com, dengan memanfaatkan peringkasan teks otomatis menggunakan metode *Latent Semantic Analysis*. Berdasarkan hasil pengujian, diperoleh rata-rata nilai *f-measure* sebesar 90,68% dan *recall* sebesar 85%, dengan rasio data latih dan data uji sebesar 90:10. Proses pengumpulan data berlangsung dari Februari hingga

Juni 2020, dengan total 200 dokumen latih dan 20 dokumen uji. Pengujian dilakukan dengan menerapkan *compression rate* sebesar 30%. [12]