

BAB I

PENDAHULUAN

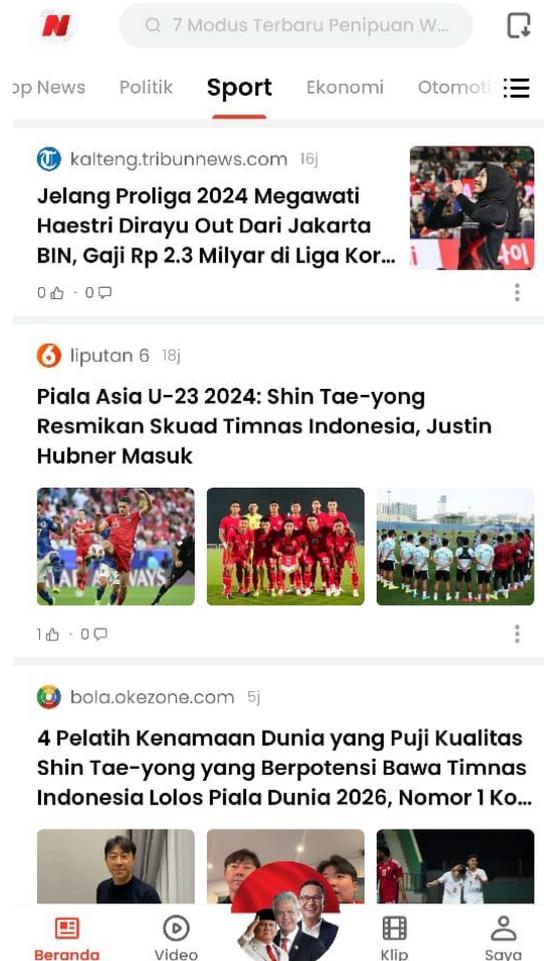
1.1 Latar Belakang

Berita telah menjadi salah satu sumber informasi yang sangat penting dalam kehidupan sehari-hari. Isi berita memainkan peran kunci dalam membentuk opini masyarakat, terutama mengenai kebijakan publik dan isu-isu politik (Ratna dkk., 2025). Berdasarkan data Dewan Pers (2024), jumlah media massa di Indonesia yang terverifikasi mencapai 1.800, dengan 1.015 diantaranya adalah media *online*. Banyaknya berita yang beredar, terutama melalui media *online*, dapat membuat masyarakat merasa kewalahan yang mendorong perilaku menghindari berita (Tian, 2022). Selain itu, berita dengan topik yang serupa sering kali diterbitkan oleh berbagai media dalam bentuk yang hampir sama. Hal ini terjadi karena beberapa media menggunakan sumber yang sama, sehingga pembaca menerima informasi yang redundan (Alonso dkk., 2013).

Sejumlah pendekatan telah dilakukan untuk meningkatkan pengalaman pengguna dalam mengakses berita secara *online*, salah satunya adalah dengan membangun platform agregasi berita. Platform agregasi berita dapat mengurangi waktu yang diperlukan untuk mengakses berbagai situs berita dalam mencari informasi (Calzada & Gil, 2020). Selain itu, platform agregasi berita juga dapat menguntungkan media berita *online*, salah satunya meningkatkan pangsa pasar (Calzada & Gil, 2020). Dengan meningkatnya kesadaran masyarakat akan konten dari media berita, jumlah pengunjung serta waktu membaca akan meningkat. Hal ini terlihat saat terjadi penutupan platform agregasi berita Google News di Spanyol, yang menyebabkan penurunan jumlah pengunjung ke situs berita di Spanyol hingga 14% (Calzada & Gil, 2020).

Di Indonesia, mayoritas masyarakat mengonsumsi berita melalui media berita *online* yaitu sebanyak 84% pada tahun 2023, disusul media sosial, televisi, dan terakhir media cetak (Annur, 2023). Terdapat beragam platform agregasi berita populer berbahasa Indonesia, seperti Opera News, CNN Indonesia, dan Google Berita. Gambar 1.1 menunjukkan contoh aplikasi agregasi berita berbahasa

Indonesia, Opera News. Di dalamnya terdapat berbagai berita yang diambil dari beberapa sumber media berita *online*, seperti tribunnews, liputan 6, dan okezone.



Gambar 1.1 Contoh Aplikasi Agregasi Berita Opera News (Sumber: Tangkapan layar aplikasi Opera News, diambil 11 April 2024)

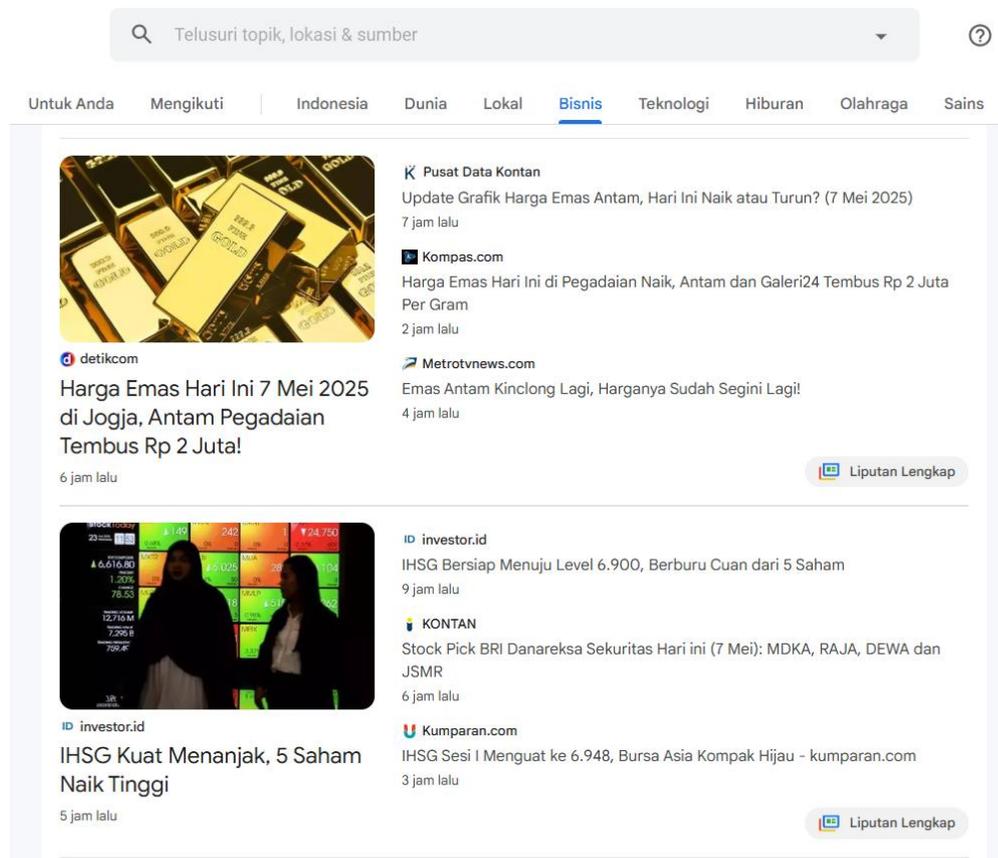
Meskipun platform agregasi berita telah terbukti dapat mengurangi waktu yang diperlukan untuk mengakses berbagai situs berita dalam mencari informasi (Calzada & Gil, 2020), kualitas konten yang disajikan juga mempengaruhi pengalaman pengguna. Salah satu tantangan utama yang dihadapi oleh platform agregasi berita adalah memastikan bahwa konten yang disajikan relevan dan bermanfaat bagi pengguna. Tanpa proses penyaringan, pengguna dapat merasa kebingungan dan terbebani oleh banyaknya informasi yang tidak relevan dengan mereka, duplikasi atau bahkan tidak akurat. Oleh karena itu, pengklasifikasian

berita yang efektif sangat penting untuk memastikan bahwa platform agregasi berita dapat menyajikan konten yang sesuai dengan minat dan preferensi pengguna.

Pendekatan yang dapat dilakukan adalah penggunaan algoritma *machine learning* untuk pengklasifikasian berita berdasarkan topik atau kategori tertentu. Dengan menggunakan teknik ini, platform agregasi berita dapat secara otomatis menyaring dan mengelompokkan berita-berita yang relevan ke dalam kategori-kategori tertentu, seperti politik, ekonomi, hiburan, olahraga, dan lain sebagainya. Penggunaan algoritma *machine learning* untuk klasifikasi berita telah mencapai akurasi 94,99% dengan *Support Vector Machine (SVM)*, 94,60% dengan *Logistic Regression*, dan 95,50% dengan *Multi-Layer dense Neural Network* pada penelitian menggunakan data berita berbahasa Bengali (Yeasmin dkk., 2021).

Selain pengklasifikasian berita berdasarkan kategori tertentu, pendekatan *clustering* serta fitur peringkasan berita juga memiliki peran yang penting dalam pengembangan platform agregasi berita yang optimal. Dengan adanya teknik *clustering*, platform agregasi berita dapat secara otomatis mengelompokkan berita-berita yang memiliki kesamaan topik atau konten, meskipun tidak ada label kategori yang tersedia. Hal ini menjadi penting karena topik-topik baru terus bermunculan setiap harinya dalam dunia berita yang terus berkembang.

Gambar 1.2 memperlihatkan contoh implementasi klasifikasi yang disertai *clustering* berita pada *website* Google Berita. Berita pertama-tama diklasifikasi berdasarkan kategori seperti bisnis, teknologi, dan olahraga. Setelah itu, berita yang sudah terbagi berdasarkan kategori, kemudian dibagi lagi berdasarkan topik yang ada pada saat itu menggunakan metode *clustering*. Topik yang sedang muncul pada kategori bisnis pada Gambar 1.2 adalah mengenai harga emas dan saham.



Gambar 1.2 Contoh Implementasi Klasifikasi dan Clustering Berita pada Website Agregasi Berita Google Berita (Sumber: Google Berita, diakses dari <https://news.google.com> pada 7 Mei 2025)

Disamping itu, peringkasan berita memungkinkan pengguna untuk mendapatkan informasi penting dari berbagai artikel berita dalam waktu singkat, tanpa harus membaca seluruh isi berita. Hal ini sangat berguna saat pengguna memiliki keterbatasan waktu atau hanya tertarik pada inti dari suatu berita. Gambar 1.3 menunjukkan contoh fitur ringkasan pada salah satu aplikasi agregasi berita, Daillyhunt. Dengan adanya fitur ini, pengguna dapat memahami inti dari suatu artikel berita tanpa harus mengunjungi sumber berita tersebut dan membaca keseluruhan teks.



Gambar 1.3 Contoh Fitur Ringkasan pada Aplikasi Agregasi Berita Dailyhunt
(Sumber: Tangkapan layar aplikasi Dailyhunt, diambil 11 April 2024)

Salah satu pendekatan yang dapat digunakan untuk melakukan peringkasan berita secara otomatis adalah melalui pemanfaatan *Large Language Model* (LLM). LLM telah terbukti mampu menghasilkan ringkasan teks yang lebih baik daripada ringkasan yang dibuat secara manual oleh manusia (Liu dkk., 2023). Keunggulan ini membuat LLM sangat berpotensi untuk diintegrasikan ke dalam sistem agregasi berita modern.

Berdasarkan uraian di atas, skripsi ini membahas pengembangan sistem agregasi berita yang dapat mengambil berita dari berbagai sumber untuk kemudian dilakukan klasifikasi serta *clustering* agar berita dapat terbagi berdasarkan topik

secara otomatis. Selain itu, sistem ini juga dilengkapi dengan fitur ringkasan otomatis berdasarkan topik yang telah dikelompokkan, guna memudahkan pengguna memahami suatu isu secara menyeluruh dalam waktu yang singkat.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan, maka akan diangkat rumusan masalah dalam penelitian ini di antaranya:

- 1) Bagaimana merancang dan menerapkan sistem untuk melakukan klasifikasi kategori berita secara otomatis?
- 2) Bagaimana melakukan *clustering* berita yang telah diklasifikasi ke dalam topik yang lebih spesifik?
- 3) Bagaimana menghasilkan ringkasan berita secara otomatis dari hasil pengelompokan berita?
- 4) Bagaimana mengintegrasikan proses klasifikasi, *clustering* dan peringkasan dalam satu sistem agregasi berita yang dapat bekerja secara otomatis?

1.3 Batasan Masalah

Dalam rangka mencapai tujuan penelitian yang telah direncanakan sebelumnya, peneliti menetapkan beberapa batasan masalah untuk penelitian ini sebagai berikut:

- 1) Data yang digunakan untuk pelatihan model klasifikasi dan peringkasan berasal dari dataset IndoSum, yang tersedia secara publik melalui platform Kaggle.
- 2) Data berita yang digunakan dalam sistem agregator berita diperoleh melalui proses web scraping dari situs media daring berbahasa Indonesia, yaitu Kompas dan TribunNews pada bulan Juni 2025.
- 3) Hanya bagian teks utama dari artikel berita yang digunakan. Elemen yang bukan teks seperti gambar, video, dan komentar pengguna tidak disertakan dalam pemrosesan.

- 4) Berita yang digunakan dalam sistem hanya mencakup kategori tajuk utama, teknologi, hiburan, olahraga, showbiz, dan inspirasi, sesuai dengan klasifikasi kategori yang ditentukan dalam pelatihan

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, maka tujuan yang akan dicapai dari penelitian ini yaitu sebagai berikut:

- 1) Merancang dan menerapkan sistem untuk melakukan klasifikasi kategori berita secara otomatis.
- 2) Melakukan *clustering* berita yang telah diklasifikasi ke dalam topik yang lebih spesifik.
- 3) Menghasilkan ringkasan berita secara otomatis dari hasil pengelompokkan berita.
- 4) Mengintegrasikan proses klasifikasi, *clustering* dan peringkasan dalam satu sistem agregasi berita yang dapat bekerja secara otomatis.

1.5 Manfaat Penulisan

Penelitian yang akan dilaksanakan ini diharapkan dapat memberikan manfaat bagi pihak-pihak yang bersangkutan, di antaranya:

- 1) Bagi Masyarakat
 - a. Dapat mengakses berita yang lebih relevan dan sesuai dengan minat mereka dengan lebih mudah dan cepat.
 - b. Kemudahan untuk selalu mendapatkan informasi terbaru dengan berita-berita terkini, sehingga dapat meningkatkan kesadaran mereka akan isu-isu penting di sekitar.
- 2) Bagi Peneliti
 - a. Peneliti akan memberikan kontribusi ilmiah dalam bidang kecerdasan buatan dan teknologi informasi dengan mengembangkan sistem agregasi berita yang memanfaatkan teknologi kecerdasan buatan.
 - b. Penelitian ini akan membuka jalan untuk pengembangan metode baru dalam pengolahan bahasa alami pada konten berita.

- c. Peneliti akan memperoleh pengalaman praktis dalam mengembangkan sistem berbasis algoritma kecerdasan buatan.
- 3) Bagi Media Berita *Online*
- a. Dengan adanya sistem agregasi berita untuk klasifikasi konten, situs berita *online* dapat mengalami peningkatan jumlah pengunjung.
 - b. Dengan menyediakan konten yang lebih relevan dan bermanfaat bagi pengguna, situs berita *online* juga dapat meningkatkan interaksi pengguna pada situs mereka.