

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi informasi dan komunikasi dalam satu dekade terakhir menyebabkan peningkatan signifikan terhadap jumlah data digital, khususnya data yang berbasis teks. Setiap hari, ribuan hingga jutaan artikel berita dipublikasikan secara daring oleh berbagai portal media nasional, sehingga masyarakat semakin bergantung pada sumber digital dalam memperoleh informasi. Kondisi seperti ini menjadikan temu balik informasi sebagai salah satu bidang yang sangat penting, terutama dalam membantu dalam menemukan konten atau informasi yang relevan dari kumpulan dokumen yang besar.

Salah satu topik berita yang mengalami peningkatan volume sejak masa pemilihan presiden tahun 2024 adalah Asta Cita. Asta Cita sendiri merupakan delapan agenda pembangunan utama yang menjadi kebijakan unggulan pada pemerintahan Prabowo-Gibran. Asta Cita sendiri sering muncul pada pemberitaan media digital seperti Kompas, CNN Indonesia, Antara News, Detik dan berbagai media lain. Pemberitaan mengenai Asta Cita sering muncul dalam berbagai konteks seperti pada konteks kebijakan ekonomi, sosial, pertahanan hingga transformasi digital. Pemberitaan tersebut tidak hanya bersifat informatif, namun juga beragam dari berbagai sudut pandang, analisis dan cakupan isu.

Tingginya perhatian publik terhadap isu asta cita dapat diamati melalui data pada *Google Trends*, yang menunjukkan adanya minat pencarian masyarakat terhadap kata kunci “Asta Cita” dengan pola yang fluktuatif dalam kurun waktu tertentu. Berdasarkan data pada *Google Trends* yang diakses pada 10 Januari 2025, dalam 30 hari terakhir kata kunci “Asta Cita” menunjukkan variasi minat pencarian antarwilayah. Beberapa contohnya yaitu DKI Jakarta tercatat tingkat minat pencarian tertinggi dengan indeks 100, diikuti oleh Jawa Tengah (83), Jawa Timur (74), Gorontalo (61), dan Banten (54). Nilai tersebut merepresentasikan tingkat popularitas relatif pencarian, bukan jumlah pencarian sebenarnya. Hal ini menunjukkan bahwa perhatian publik terhadap isu Asta Cita tersebar di berbagai wilayah dengan minat yang berbeda-beda.

Pemberitaan media Kompas dalam artikelnya yang berjudul “Asta Cita, Visi Pemerintahan Prabowo-Gibran Menuju Indonesia Emas 2045”, Asta Cita dirancang sebagai arah pembangunan nasional yang mencakup fokus utama, mulai dari percepatan swasembada pangan, pemerataan ekonomi, percepatan digitalisasi, peningkatan kualitas kesehatan, hingga penguatan pertahanan negara. Kompas menjelaskan bahwa Asta Cita tidak hanya menjadi visi politik, namun juga fondasi kebijakan negara menuju Indonesia Emas 2045, sehingga cakupan informasinya sangat luas dan diberitakan dari banyak sudut pandang[1].

Selain itu, pemberitaan oleh Antara News juga menunjukkan bahwa Asta Cita memperoleh perhatian dalam konteks pembangunan sumber daya manusia lintas generasi. Dalam artikelnya yang berjudul “Peringatan Hari Kartini : Wujudkan Asta Cita untuk Indonesia Emas 2025”, Antara News menyampaikan bahwa implementasi Asta Cita dianggap sangat penting untuk mendukung kolaborasi antara generasi muda dan perempuan dalam pembangunan nasional. Hal ini memperkuat bahwa isu Asta Cita tidak hanya bersifat politis, tapi juga sosial dan kultural, sehingga cakupan pemberitaannya menjadi semakin luas[2].

Selanjutnya, pada pemberitaan dalam artikel Detik hikmah menyoroti aspek dukungan publik terhadap Asta Cita, terutama dalam peningkatan kualitas SDM. Dalam artikelnya yang berjudul “Dukung Asta Cita, NU Dorong Pengembangan SDM-Kesejahteraan Perempuan”, dijelaskan bahwa berbagai organisasi masyarakat, lembaga keagamaan, dan komunitas pemuda memberi perhatian khusus dalam implementasi program-program Asta Cita. Artikel tersebut memperlihatkan bahwa perhatian masyarakat terhadap Asta Cita melahirkan beragam opini, analisis kebijakan, serta narasi nasional yang semakin memperkaya data berita yang beredar di media digital[3].

Variasi mengenai penulisan, istilah, sudut pandang serta konteks pemberitaan yang menjadikan informasi mengenai Asta Cita menjadi kompleks. Dengan keberagaman tersebut menimbulkan tantangan dalam proses pencarian informasi. Oleh karena itu, kualitas dari temu balik informasi (*Information Retrieval*) menjadi sangat penting dalam memastikan bahwa dokumen yang ditampilkan dapat benar-benar sesuai dengan kebutuhan informasi pengguna yang mencari informasi terkait Asta Cita.

Dalam bidang sistem temu kembali informasi, representasi teks berperan dalam menentukan kualitas pencarian dokumen. Salah satu metode klasik yang sering digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). Model ini memberikan bobot pada kata dasar berdasarkan frekuensi kemunculan dalam sebuah dokumen dan pada seluruh korpus, sehingga TF-IDF mampu menonjolkan kata-kata yang dianggap penting. Salah satu contohnya pada penelitian yang dilakukan oleh Putu et al. (2025), TF-IDF digunakan bersama dengan *cosine similarity* untuk menilai kemiripan antara kueri pengguna dengan dokumen jawaban, serta berhasil memperoleh hasil *retrieval* dengan akurasi yang tinggi[4].

Meskipun demikian, metode TF-IDF memiliki keterbatasan yang signifikan yaitu metode ini tidak memasukkan informasi semantik seperti makna kata, sinonim, maupun konteks kalimat dalam perhitungannya. TF-IDF memperlakukan kata sebagai unit yang berdiri sendiri sehingga meskipun terdapat dua dokumen yang memiliki konsep serupa, sistem dapat gagal mengenali kecocokan jika tidak ada kesamaan kata. Hal ini dapat menyebabkan kegagalan dalam menemukan dokumen yang relevan ketika istilah yang digunakan berbeda namun memiliki makna yang sama.

Untuk mengatasi keterbatasan dari metode tersebut, model representasi teks yang modern seperti *word embedding* mulai digunakan secara luas. Salah satu metode *word embedding* yaitu *Word2Vec*, dimana metode ini mampu memetakan kata dalam ruang vektor yang berdimensi rendah sehingga kata-kata dengan konteks serupa akan memiliki representasi berdekatan. Pada penelitian yang dilakukan Suryaningrum dkk, yang menunjukkan bahwa *Word2Vec* dapat menghasilkan performa yang lebih baik dibandingkan dengan model TF-IDF karena model *Word2Vec* mampu mempresentasikan hubungan semantik antar kata sehingga mendukung dalam peningkatan akurasi klasifikasi dan pencarian dokumen[5].

Perkembangan *Natural Language Processing* (NLP) kemudian menghasilkan model berbasis *transformer* seperti *Bidirectional Encoder Representations from Transformers* (BERT). Salah satu turunan dari BERT ini adalah IndoBERT yang dikembangkan khusus untuk teks berbahasa Indonesia.

Model ini mampu membaca konteks kata secara *bidirectional* dan menghasilkan representasi yang lebih akurat. Seperti pada penelitian yang dilakukan oleh Sunendar & Saputra, menunjukkan bahwa IndoBERT terbukti memberikan performa temu balik informasi yang lebih tinggi dibanding dengan model berbasis LTSM[6]. Hasil ini mengindikasikan bahwa model berbasis *transformer* seperti IndoBERT mampu memahami konteks semantik dokumen secara lebih mendalam, sehingga menghasilkan pencarian yang lebih relevan dan akurat dibandingkan metode representasi tradisional. Oleh karena itu, model ini dapat diadaptasi dalam penelitian temu balik informasi.

Berdasarkan masalah, kebutuhan informasi, serta temuan dari penelitian terdahulu maka pada penelitian ini berfokus pada pengembangan sistem temu kembali informasi pada korpus berita Asta Cita melalui perbandingan model representasi teks. Untuk mencapai hal tersebut, maka dalam penelitian ini membandingkan tiga pendekatan representasi teks yang memiliki karakteristik berbeda, yaitu TF-IDF sebagai model klasik berbasis frekuensi, *Word2Vec* sebagai model representasi vektor yang mampu menangkap hubungan semantik antar kata, serta IndoBERT sebagai model berbasis *transformer* yang mampu memahami konteks secara *bidirectional* pada teks berbahasa Indonesia. Berdasarkan hasil evaluasi kinerja model dengan hasil terbaik tersebut yang akan digunakan sebagai dasar untuk membangun sistem temu balik informasi.

Oleh karena itu, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan temu balik informasi khususnya pada domain berita berbahasa Indonesia yang membutuhkan model representasi teks yang mampu menangani konteks semantik dengan lebih baik. Hasil dari penelitian ini juga diharapkan dapat menjadi dasar dalam pengembangan sistem IR yang lebih efektif, akurat, serta adaptif dalam mendukung pencarian informasi masyarakat dalam era digital.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan dalam penelitian ini, didapatkan rumusan masalah sebagai berikut ini :

1. Bagaimana perbandingan kinerja dari model representasi teks TF-IDF, *Word2Vec* dan IndoBERT dalam proses temu balik informasi pada berita Asta Cita?
2. Bagaimana menentukan model dengan tingkat relevansi hasil pencarian dokumen terbaik berdasarkan metrik evaluasi serta mengimplementasikannya ke dalam sistem temu balik informasi Asta Cita?

1.3. Batasan Masalah

Untuk memperjelas fokus dari penelitian, berikut ini batasan-batasan yang ditetapkan untuk membatasi ruang lingkup yang akan dibahas pada penelitian ini :

1. Data pada penelitian ini berupa berita daring yang membahas mengenai Asta Cita pada masa pemerintahan Prabowo-Gibran dari portal berita nasional.
2. Penelitian ini berfokus pada representasi teks dan temu balik informasi
3. Model representasi yang digunakan dibatasi pada TF-IDF, *Word2Vec* dan IndoBERT.
4. Proses pengambilan data ini dilakukan melalui proses *web scraping* menggunakan *Google Colab* dengan bantuan *API news search* untuk memastikan data yang diperoleh sesuai dengan kata kunci.
5. Sistem akan dibangun hanya sampai pada tahap aplikasi yang berfungsi untuk memvisualisasikan hasil dari proses *information retrieval*, berupa: *insight berita*, halaman pencarian, halaman hasil pencarian, dan *home* untuk menampilkan informasi tentang Asta Cita.

1.4. Tujuan Penelitian

Adapun tujuan dari penelitian ini dirumuskan berdasarkan latar belakang adalah sebagai berikut :

1. Menerapkan dan membandingkan model representasi teks, yaitu, TF-IDF, *Word2Vec*, dan IndoBERT pada temu balik informasi berita Asta Cita.
2. Menentukan model dengan tingkat relevansi pencarian dokumen terbaik berdasarkan metrik evaluasi serta mengimplementasikannya dalam sistem temu balik informasi Asta Cita.

1.5. Manfaat Penelitian

Mengacu pada tujuan yang telah dijelaskan sebelumnya maka penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Memberikan kontribusi bagi pengembangan penelitian pada bidang *Natural Language Processing* (NLP) dan *Information Retrieval* (IR), khususnya terkait penerapan dan perbandingan representasi teks yaitu TF-IDF, *Word2Vec*, dan IndoBERT pada temu balik informasi.
2. Penelitian ini diharapkan dapat memberikan gambaran mengenai model representasi teks yang memiliki performa terbaik yang dijadikan pertimbangan dalam pengembangan sistem temu balik informasi berbasis teks berbahasa Indonesia.

1.6. Sistematika Penulisan

Sistematika penulisan skripsi ini bertujuan untuk memandu penyusunan laporan dan menjadi pedoman dalam mencapai tujuan penulisan skripsi. Berikut merupakan bagian yang digunakan dalam proses penyusunan skripsi:

BAB I PENDAHULUAN

Bab ini menyajikan gambaran umum terkait isi penelitian skripsi yang akan dilakukan. Bab pendahuluan berisikan latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat penelitian, dan sistematika penelitian yang digunakan.

BAB II TINJAUAN PUSTAKA

Bab berisikan pengertian umum maupun khusus sebagai teori-teori dan penelitian-penelitian terdahulu yang relevan berkaitan dengan topik penelitian. Serta dalam tinjauan pustaka ini memberikan landasan teoritis yang kuat untuk penelitian yang akan dilakukan.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah yang diambil untuk mencapai tujuan penelitian, mencakup identifikasi masalah, studi literatur, metode pengumpulan data, analisis kebutuhan, dan desain sistem.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan pengujian metode temu balik informasi sesuai tahapan dalam metodologi penelitian, yang mencakup pengolahan data, penerapan metode representasi teks, serta evaluasi kinerja pencarian.

BAB V PENUTUP

Bab ini memuat ringkasan dari keseluruhan hasil penelitian yang telah dibahas sebelumnya, serta menyajikan saran sebagai bahan pertimbangan untuk pengembangan penelitian selanjutnya.

DAFTAR PUSTAKA

Bagian ini mencantumkan literatur yang digunakan sebagai referensi dalam penyusunan skripsi, memberikan dasar yang kuat bagi argumen dan analisis yang disajikan.

LAMPIRAN

Pada bagian lampiran, memuat berbagai data serta informasi yang pelengkap yang mendukung dalam penyusunan proposal skripsi ini.