

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saat ini informasi telah menjadi kebutuhan penting dalam kehidupan. Salah satu sumber informasi utama adalah berita, yaitu informasi terkini tentang suatu peristiwa yang terjadi. Sebelum era digital, kebutuhan berita dipenuhi dari berbagai pemberitaan oleh media masa cetak, televisi, maupun radio. Seiring perkembangan teknologi informasi yang demikian pesat, era digital telah memanjakan manusia dengan segala kemudahannya untuk mendapatkan berbagai informasi yang diinginkan. Salah satu kemudahan tersebut adalah dengan munculnya berbagai portal berita *online*[1].

Berita merupakan sarana yang digunakan oleh masyarakat untuk memperoleh informasi mengenai suatu peristiwa atau kejadian *faktual* secara tertulis. Teknologi yang semakin berkembang menjadikan berita lebih mudah untuk dijangkau oleh masyarakat, salah satunya dengan adanya portal berita *online*. Banyaknya jumlah portal berita *online* yang ada maka artikel berita yang beredar juga semakin banyak. Data berita terus bertambah seiring berjalannya waktu sehingga menjadi tantangan tersendiri untuk mengolah data berita menjadi bentuk yang lebih bermanfaat yaitu dengan menemukan topik tersembunyi dari sekumpulan berita melalui pemodelan topik. Menemukan topik dari sekumpulan berita, dapat menghemat lebih banyak waktu daripada membaca semuanya untuk mengetahui peristiwa yang terjadi dalam jangka waktu tertentu[2].

Berbagai informasi terdapat pada portal berita baik dari daerah, dalam dan luar negeri, serta beraneka ragam subjek ulasan. Data berita pada portal berita *online* selain sebagai sumber informasi juga dapat digunakan sebagai bahan analisis serta implementasi penelitian. Dari berbagai macam kategori berita yang dibahas dalam portal berita memiliki topik-topik tertentu, namun proses identifikasi memerlukan waktu yang lama untuk menentukan topik mana yang menjadi trend dalam suatu portal berita. Oleh karena itu, pada penelitian ini diharapkan dapat memberikan solusi untuk masalah identifikasi topik pada artikel berita yang diterbitkan oleh situs berita *online* yaitu *Detik.com*[3].

Detik.com merupakan pelopor situs berita *online* yang mandiri, dimana mampu berdiri tanpa adanya media cetak sebagai pendukungnya. Selain itu, *Detik.com* juga mampu menyodorkan gaya baru yang lebih ringkas (*to the point*) dalam memuat berita yang dikemas secara *update* dan *realtime* pada situs *website*[4]. Aplikasi *mobile Detik.com* memiliki fungsi yang sangat bermanfaat bagi penggunaannya dalam mengakses berita dan informasi terkini. Dengan aplikasi ini, pengguna dapat dengan mudah mendapatkan berita terbaru dari berbagai kategori seperti politik, ekonomi, teknologi, olahraga, dan hiburan, semuanya dalam satu platform yang terintegrasi. Aplikasi ini dirancang untuk memberikan pengalaman yang cepat dan efisien, memungkinkan pengguna untuk membaca berita dimana saja dan kapan saja, tanpa perlu membuka *browser*[5].

Era sekarang merupakan era di mana akses ke berita dan informasi begitu mudah, tantangan utama yang timbul adalah bagaimana mengelompokkan dan mengkategorikan berita dengan cepat dan akurat. Dengan produksi berita yang semakin besar dan beragam, memahami topik berita menjadi sangat penting untuk penyampaian informasi yang efektif kepada masyarakat[6].

Pemodelan topik merupakan salah satu teknik *unsupervised machine learning* untuk menemukan tema tersembunyi dari kumpulan dokumen teks besar guna mengelompokkan tema-tema tersebut menjadi satu topik. Pemodelan topik mempunyai banyak metode yang dapat digunakan seperti *Latent Semantic Analysis (LSA)*, *Probabilistic Latent Semantic Analysis (PLSA)*, *Latent Dirichlet Allocation (LDA)*. Pertama kali diperkenalkan oleh David Blei, *Latent Dirichlet Allocation (LDA)* merupakan salah satu pemodelan topik yang paling populer, *LDA* digunakan untuk menampilkan sebuah topik menggunakan probabilitas dari setiap kata sehingga dapat membantu menggambarkan dokumen-dokumen menjadi lebih terstruktur. Penerapan Pemodelan topik menggunakan *LDA* dianggap lebih unggul dalam menghasilkan topik bermakna logis, kemampuan interpretasi, dan performansi prediktif[7].

Pemodelan topik berbeda dengan *clustering* karena meskipun keduanya merupakan metode *unsupervised learning*, pendekatan dan tujuan utamanya tidak sama. Pemodelan topik, seperti *Latent Dirichlet Allocation (LDA)*, dirancang untuk mengungkap struktur laten berupa topik-topik tersembunyi dalam dokumen, di

mana satu dokumen dapat merepresentasikan campuran beberapa topik dengan proporsi tertentu. Ini berarti pemodelan topik mempertimbangkan distribusi kata dan makna *semantik* yang terkandung dalam teks. Sementara itu, *clustering* lebih menekankan pada pengelompokan dokumen berdasarkan kemiripan fitur numerik, seperti jarak vektor *TF-IDF*, tanpa memahami konteks *semantik* atau struktur topik di dalam teks. *Clustering* juga bersifat hard assignment, yaitu setiap dokumen hanya masuk ke dalam satu cluster, berbeda dengan pemodelan topik yang bersifat yakni setiap elemen bisa berhubungan dengan beberapa kelompok sekaligus.

Pemodelan topik, dapat diimplementasikan metode pengelompokan berdasarkan ukuran kedekatan (kemiripan) suatu data, salah satunya dengan menggunakan *Latent Dirichlet Allocation (LDA)*. Metode ini merupakan sebuah metode *text mining* untuk menemukan pola tertentu pada sebuah dokumen dengan menghasilkan beberapa macam topik yang berbeda, sehingga tidak secara spesifik mengelompokkan dokumen[8]. Metode *LDA* akan menemukan kata yang ada pada topik secara semi acak distribusi, menghitung probabilitas topik pada dataset, dan menghitung probabilitas terhadap topik setiap iterasinya[9]. sehingga menghasilkan pemodelan topik yang lebih terstruktur dan relevan. Proses ini dilakukan dengan mendeteksi pola kata yang sering muncul bersama dalam dokumen tertentu, kemudian membaginya ke dalam kelompok topik yang berbeda. Dengan demikian, *LDA* mampu memberikan wawasan yang mendalam tentang struktur tema yang mendominasi dalam kumpulan data teks, yang sangat bermanfaat untuk analisis lebih lanjut seperti pengelompokan dokumen, pengambilan informasi, atau visualisasi hubungan antar topik.

Penelitian berjudul *Identifikasi Topik Artikel Berita menggunakan Topic modelling dengan Latent Dirichlet Allocation*, Penelitian ini mengembangkan sistem identifikasi topik artikel berita kriminal menggunakan *Latent Dirichlet Allocation (LDA)*. Data diperoleh melalui **web scraping** dari **Detik.com** dan **Tempo.co**, dengan total 20 artikel. Setelah melalui proses pembersihan dan *preprocessing* (*Tokenization*, *Case folding* , *Stopword removal*, serta *Stemming/lemmatization*), *LDA* diterapkan dengan variasi jumlah topik (2, 3, 4, 5, dan 7) dan iterasi (500, 1000, dan 5000). Hasil terbaik diperoleh pada **jumlah** topik 3, dengan klaster yang tidak saling beririsan. Hasil penelitian menunjukkan bahwa

LDA efektif dalam mengelompokkan artikel berdasarkan pola tematik yang muncul dalam teks. Namun, penelitian ini masih memiliki keterbatasan dalam hal analisis temporal karena tidak mengeksplorasi bagaimana tren topik berita berkembang dari waktu ke waktu. Selain itu, penelitian ini hanya berfokus pada identifikasi topik secara statis tanpa mempertimbangkan dampak peristiwa politik terhadap perubahan topik berita. Oleh karena itu, diperlukan penelitian yang lebih mendalam untuk memahami perubahan tren topik berita politik dalam jangka waktu tertentu guna memberikan wawasan mengenai pola isu yang berkembang[10].

Penelitian berjudul *Topic modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM*, Penelitian ini menggunakan metode *Long Short-term Memory (LSTM)* dengan memanfaatkan *pre-trained global vector for word representation* berbahasa Indonesia yang diperoleh dari fasttext menghasilkan model dengan nilai akurasi tertinggi sebesar 71.13% dan nilai akurasi terendah sebesar 63.48%. Pada setiap eksperimen hyperparameters tuning, proses ini menghasilkan nilai akurasi yang berbeda beda sesuai dengan parameter-parameter yang dipasang, yaitu jumlah lapisan *LSTM*, jumlah *recurrent units*, *dropouts*, *ukuran batch*, dan *maksimum epoch*. Pemodelan topik pada penelitian ini menggunakan metode *Latent Dirichlet Allocation (LDA)* dapat menghasilkan topik-topik yang dibicarakan oleh media *online*, baik itu topik topik yang bersentimen positif, negatif, maupun netral. Namun, penelitian ini lebih berfokus pada analisis sentimen dalam judul berita dan tidak secara spesifik mengeksplorasi tren topik berita politik dalam artikel lengkap. Selain itu, studi ini tidak membahas pola perubahan topik dalam jangka waktu tertentu, yang menjadi aspek penting dalam memahami dinamika politik. Oleh karena itu, penelitian lebih lanjut diperlukan untuk menganalisis tren topik politik secara *longitudinal* guna melihat pola perubahan isu yang berkembang di media berita *online*[11].

Penelitian berjudul *Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA)*, Penelitian ini menganalisis topik informasi publik yang tersebar melalui media sosial Radio Suara Surabaya menggunakan metode *Latent Dirichlet Allocation (LDA)*. Dengan meningkatnya jumlah laporan harian yang diterima dari masyarakat melalui media sosial, dibutuhkan sistem otomatis untuk mengklasifikasikan topik-topik yang

muncul. *LDA* sebagai metode *text mining*, digunakan untuk mengidentifikasi pola dalam dokumen dan mengelompokkan laporan berdasarkan topik tertentu. Hasil eksperimen menunjukkan bahwa terdapat empat topik utama dalam pesan media sosial yang dianalisis, dengan nilai *perplexity* terbaik sebesar 213,41. Model yang dihasilkan juga telah diuji melalui uji koherensi topik menggunakan word intrusion task dan topic intrusion task, yang menunjukkan bahwa topik yang diperoleh dapat diinterpretasikan dengan baik oleh manusia. Penelitian ini menunjukkan bahwa *LDA* dapat digunakan secara efektif untuk menemukan pola komunikasi publik terkait berbagai isu di platform digital. Namun, penelitian ini hanya berfokus pada informasi publik di media sosial dalam lingkup lokal, bukan pada analisis berita politik dalam skala nasional. Selain itu, data yang digunakan dalam penelitian ini berasal dari media sosial, yang memiliki karakteristik berbeda dibandingkan artikel berita yang lebih terstruktur. Oleh karena itu, masih diperlukan penelitian yang lebih komprehensif untuk menganalisis trending topik berita politik berbasis artikel jurnalistik yang diterbitkan oleh media berita besar seperti *Detik.com* guna memahami pola pemberitaan politik secara lebih sistematis[12].

Berdasarkan penelitian-penelitian sebelumnya, masih terdapat beberapa celah penelitian (*research gap*) yang dapat dieksplorasi lebih lanjut. Sebagian besar penelitian yang menggunakan *Latent Dirichlet Allocation (LDA)* hanya berfokus pada identifikasi topik dalam satu periode waktu tertentu tanpa mempertimbangkan perubahan tren topik dalam jangka waktu yang lebih panjang. Selain itu, beberapa studi lebih menitikberatkan pada analisis sentimen atau informasi publik di media sosial, yang memiliki karakteristik berbeda dibandingkan dengan artikel berita politik yang lebih terstruktur. Metode yang digunakan dalam penelitian sebelumnya juga cenderung belum mempertimbangkan aspek temporal dalam analisis topik politik, sehingga pola perubahan isu politik dari waktu ke waktu masih belum tergambar secara jelas.

Penelitian ini bertujuan untuk melengkapi kekurangan tersebut dengan menganalisis trending topik dalam berita politik yang dipublikasikan di *Detik.com* dalam rentang waktu tertentu menggunakan metode *LDA*. Penelitian ini akan mengeksplorasi bagaimana topik-topik politik berkembang dari waktu ke waktu dengan mengelompokkan berita berdasarkan bulan publikasi, sehingga dapat

memberikan pemahaman yang lebih mendalam mengenai dinamika isu politik yang muncul di media. Dengan pendekatan ini, penelitian diharapkan dapat mengidentifikasi pola perubahan topik politik serta mengevaluasi efektivitas metode *LDA* dalam menangkap tren berita politik. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pemodelan topik berita politik secara lebih akurat, serta membantu pengembangan sistem analisis berita yang lebih canggih dalam memahami isu-isu politik yang berkembang di masyarakat.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Seiring dengan meningkatnya jumlah artikel berita yang diterbitkan secara *online*, kebutuhan untuk mengelompokkan dan memahami topik dari kumpulan berita menjadi semakin penting. Namun, proses identifikasi topik secara manual memiliki beberapa tantangan. Pertama, waktu yang diperlukan untuk membaca dan memahami seluruh artikel sangat lama. Kedua, banyaknya data berita yang terus bertambah menyebabkan kesulitan dalam menganalisis tren topik terkini. Ketiga, tidak adanya sistem otomatis yang dapat membantu mengidentifikasi topik dari kumpulan berita secara cepat dan akurat semakin memperbesar tantangan ini.

1.2.2 Metode Usulan

Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan penggunaan metode *Latent Dirichlet Allocation (LDA)*, salah satu teknik pemodelan topik yang populer dalam *text mining*. Metode *LDA* digunakan untuk mengelompokkan artikel berita berdasarkan topik-topik yang ditemukan secara otomatis. Selain itu, *LDA* dapat mengidentifikasi distribusi probabilitas kata dalam setiap topik sehingga memberikan interpretasi yang bermakna. Dengan metode ini, diharapkan dapat dihasilkan analisis topik yang berguna untuk memahami tren berita terkini secara lebih efisien.

1.2.3 Pertanyaan Penelitian

Penelitian ini berfokus pada tiga pertanyaan utama yakni :

1. Seberapa efektif *LDA* dalam mengidentifikasi dan mengelompokkan topik utama yang sering muncul dalam artikel berita politik di *Detik.com*?
2. Bagaimana pengaruh parameter seperti jumlah topik dan iterasi terhadap kualitas pemodelan topik menggunakan *LDA*?

3. Bagaimana analisis tren topik berdasarkan resolusi waktu (bulanan, kuartalan, tahunan) dapat memberikan wawasan tentang perkembangan isu politik?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Berdasarkan pertanyaan tersebut maka maksud dan tujuan penulisan penelitian ini adalah:

1. Mengidentifikasi dan mengelompokkan topik utama dari artikel berita politik di *Detik.com* menggunakan metode *Latent Dirichlet Allocation (LDA)*.
2. Menganalisis pengaruh parameter pemodelan seperti jumlah topik, iterasi, dan *preprocessing* teks terhadap kualitas topik yang dihasilkan.
3. Menentukan resolusi waktu terbaik dalam analisis tren topik berita politik untuk memahami pola perubahan isu dari waktu ke waktu.

1.3.2 Manfaat Penelitian

1. Bagi Peneliti: Memberikan pemahaman mendalam tentang penerapan *LDA* untuk pemodelan topik pada artikel berita *Detik.com*.
2. Bagi Pembaca: Memudahkan pembaca dalam memahami tentang pemodelan topik menggunakan metode *LDA*.
3. Bagi Penelitian Selanjutnya: Menjadi referensi untuk pengembangan penelitian lebih lanjut terkait pemodelan topik menggunakan metode *LDA*.

1.4 Batasan Masalah

Batasan-batasan dalam penelitian ini diberikan agar permasalahan tidak terlalu luas dan tetap sesuai dengan tujuan penulis, sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah artikel berita kategori politik dari portal berita *online Detik.com*.
2. Metode yang digunakan untuk pemodelan topik adalah *Latent Dirichlet Allocation (LDA)*.
3. Penelitian hanya fokus pada pemodelan topik dan analisis trend.

1.5 Sistematika Penulisan

Sistematika penulisan proposal skripsi ini terdiri dari beberapa bagian sebagai berikut:

BAB I: PENDAHULUAN

Bab ini menguraikan latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta sistematika penulisan sebagai panduan struktur keseluruhan skripsi.

BAB II: TINJAUAN PUSTAKA

Berisi tinjauan literatur terkait konsep dan teori yang relevan, terutama mengenai analisis sentimen, metode *LDA*. Juga mencakup penelitian terdahulu yang mendukung teori dan metodologi penelitian.

BAB III: METODE USULAN

Bab ini menjelaskan metodologi penelitian secara rinci, termasuk pengumpulan data ulasan, perancangan model menggunakan *LDA*, serta pengujian dan evaluasi model. Tujuan utamanya adalah mendokumentasikan langkah-langkah penelitian, memastikan validitas, serta memungkinkan replikasi oleh peneliti lain.

BAB IV: HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil dari penerapan metode *Latent Dirichlet Allocation (LDA)* dalam menganalisis artikel berita politik dari portal *Detik.com* serta pembahasan atas temuan-temuan yang diperoleh. Analisis dimulai dari proses *preprocessing* data teks, pemodelan topik, evaluasi koherensi model, hingga visualisasi dan interpretasi topik-topik yang dihasilkan.

BAB V: PENUTUP

Bab ini menjelaskan kesimpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan dan penelitian lanjutan. Kesimpulan disusun berdasarkan hasil analisis topik dan visualisasi tren yang diperoleh dari proses pemodelan dengan *LDA*. Selain itu, bab ini juga memberikan masukan terkait potensi perbaikan metode atau pendekatan lain yang dapat digunakan dalam penelitian sejenis di masa mendatang.