

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Berita adalah informasi berupa fakta atau ide yang sedang terjadi yang bersifat menarik atau penting bagi sebagian besar masyarakat yang disajikan dalam bentuk media cetak, televisi, atau media internet [1]. Seiring dengan perkembangan teknologi, media online saat ini memanfaatkan penggunaan media sosial untuk memperluas jangkauan pembaca yang lebih memilih untuk membaca berita sambil berselancar di media sosial. Selain itu, penyajian berita menggunakan media sosial juga dapat meningkatkan partisipasi pembaca dalam membuat opini sehingga dapat mempercepat penyebaran informasi [2]. Salah satu media sosial populer yang digunakan oleh media online adalah Twitter.

Twitter merupakan salah satu media sosial yang banyak digunakan masyarakat sebagai media untuk berkomunikasi dan mendapatkan informasi. Indonesia menempati urutan ke-5 dengan jumlah pengguna Twitter terbanyak di dunia sebanyak 18,45 juta pengguna pada Januari 2022 [3]. Media online menggunakan Twitter dengan cara menulis *tweet* dan *link* yang mengarahkan ke artikel pada portal berita berbentuk web. *Tweet* berita dibuat secara otomatis ke dalam akun Twitter menggunakan *Content Management System* berisi potongan kalimat paragraf dalam isi artikel atau *headline* berita [4]. Kategori berita yang dimuat ke dalam sebuah *tweet* juga beraneka ragam seperti kejahatan, olahraga, kesehatan, dan lain-lain.

Salah satu kategori berita yang gencar diberitakan hingga saat ini adalah covid-19. Berita covid-19 dibuat dengan beberapa bentuk seperti edukasi, sosialisasi, *update* jumlah kasus, dan lain-lain. Jumlah *tweet* yang tinggi setiap harinya memunculkan berbagai topik berita terkait covid-19 yang membuat peneliti ingin mengetahui apa saja topik yang muncul. Akan tetapi, Twitter tidak menyediakan fitur pengkategorian *tweet* berdasarkan topik yang dibahas sehingga *tweet* akan tercampur dan tersusun secara urut berdasarkan *timeline*-nya [5]. Maka dari itu, diperlukan sebuah pemodelan topik untuk menentukan topik apa saja yang sedang diberitakan terkait covid-19 oleh media online.

Pemodelan topik merupakan sebuah teknik penganalisisan teks berupa algoritma untuk mengidentifikasi pola terhadap rangkaian kata dengan menggunakan

pendistribusian kata di dalam sebuah kumpulan dokumen, mempunyai beberapa metode seperti *Latent Semantic Analysis* (LSA), *Probabilistic Latent Semantic Analysis* (PLSA), dan *Latent Dirichlet Allocation* (LDA). LSA merupakan metode *Natural Language Processing* (NLP) dengan cara membuat vektor dari representasi teks untuk membuat *semantic content*. PLSA merupakan metode pendekatan untuk memperbaiki beberapa kerugian yang ada di dalam metode LSA [6]. Metode yang akan digunakan dalam penelitian ini adalah *Latent Dirichlet Allocation* (LDA). Peneliti memilih metode LDA karena merupakan metode penyempumaan dari metode LSA dan PLSA serta di dalamnya terdapat peningkatan cara model campuran dalam menangkap pertukaran kata-kata dan dokumen [7]. Metode LDA bekerja dengan cara menemukan pola tertentu di dalam sebuah dokumen yang akan menghasilkan beberapa macam topik yang berbeda [8].

Pada penelitian ini peneliti akan melakukan pemodelan topik berita pada *tweet* media online di Twitter dengan kata kunci covid-19 menggunakan metode *Latent Dirichlet Allocation* (LDA). Hasil dari penelitian ini diharapkan dapat memberikan gambaran apa saja topik-topik berita terkait covid-19 yang banyak digunakan oleh media online di Twitter dalam rentang waktu tertentu.

### 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka diperoleh rumusan masalah yang akan diteliti sebagai berikut:

1. Apa saja topik yang muncul dalam penerapan *Latent Dirichlet Allocation* (LDA) terhadap *tweet* media online di Twitter dengan kata kunci 'covid'?
2. Apakah hasil topik yang dilakukan dapat diinterpretasi oleh manusia dengan baik?
3. Berapa nilai optimal pada evaluasi kualitas model dengan *topic coherence* yang bisa didapatkan untuk menentukan jumlah topik dalam model LDA?

### 1.3 Batasan Masalah

Batasan masalah yang menjadi ruang lingkup permasalahan pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah *tweet* berita dari beberapa media online di Indonesia pada media sosial Twitter dengan kata kunci pencarian 'covid' dalam rentang waktu Oktober 2021 hingga Mei 2023.

2. Pengambilan data menggunakan Application Programming Interface (API) resmi versi gratis yang disediakan oleh Twitter.
3. Jenis data yang dibuat pemodelan topik berupa data teks.
4. Jumlah iterasi yang dilakukan untuk menghitung *coherence score* tertinggi sebagai jumlah topik model dibatasi sebanyak 15 iterasi.
5. Output dari program yang dibuat dapat digunakan untuk menganalisis topik tweet berita covid-19 yang banyak dimuat oleh media online di Twitter.

#### 1.4 Maksud dan Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka diperoleh maksud dan tujuan penelitian yang akan diteliti sebagai berikut:

1. Mengetahui topik apa saja yang muncul dalam penerapan *Latent Dirichlet Allocation* (LDA) terhadap *tweet* media online di Twitter dengan kata kunci 'covid'.
2. Mengetahui apakah hasil topik yang dilakukan dapat diinterpretasi oleh manusia dengan baik.
3. Memperoleh nilai optimal pada evaluasi kualitas model dengan *topic coherence* yang bisa didapatkan untuk menentukan jumlah topik dalam model LDA.

#### 1.5 Manfaat Penelitian

Melalui penelitian ini diharapkan dapat memberikan beberapa manfaat antara lain:

1. Memahami metode pemodelan topik menggunakan algoritma *Latent Dirichlet Allocation* (LDA) dan menerapkan pada studi kasus dalam kehidupan sehari-hari.
2. Sistem yang mempermudah menemukan topik berita terkait covid-19 yang sering diberitakan oleh media online di Twitter.
3. Masyarakat dapat terbantu dalam menerima informasi yang lebih ringkas terhadap pemberitaan covid-19 oleh media online di Twitter.

#### 1.6 Metode Penelitian

Pada metode penelitian ini akan diuraikan tahapan yang dilakukan pada proses penelitian ini agar berjalan dengan baik.

### 1.6.1 Pengumpulan data

Mencari jurnal terkait kemudian melakukan pengambilan data menggunakan *Application Programming Interface (API)* yang disediakan oleh Twitter dengan menggunakan pencarian *tweet* berdasarkan kata pencarian yaitu covid. *Tweet* diambil dari 16 media online resmi yang menggunakan bahasa Indonesia.

### 1.6.2 Preprocessing

*Preprocessing* yang dilakukan pada penelitian ini digunakan untuk membersihkan data teks seperti menghilangkan karakter yang tidak digunakan, mengubah karakter menjadi *lower case*, *stemming*, dan lain-lain. Dengan tahap *preprocessing* data diharapkan hasil yang akan diperoleh menjadi lebih akurat.

### 1.6.3 Analisis data

Analisis data dilakukan terhadap data teks yang sudah diambil sebelumnya yang masih berbentuk tidak terstruktur lalu diubah ke dalam bentuk terstruktur sehingga dapat dianalisis dan menghasilkan kesimpulan yang sesuai harapan.

### 1.6.4 Pengembangan

Setelah menganalisis data maka tahap selanjutnya akan dilakukan proses pengembangan. Pada tahap pengembangan ini akan dilakukan pembuatan sistem dengan menggunakan *tools* yang diperlukan. Pengembangan sistem dilakukan dengan cara eksperimen untuk menemukan hasil yang terbaik pada sistem yang dibuat.

### 1.6.5 Analisis hasil

Tahap selanjutnya adalah proses analisis hasil. Tahap ini diperlukan untuk mengetahui bahwa pemodelan topik berita covid-19 yang terbentuk dapat dijalankan sesuai dengan tujuan dan menghasilkan *output* yang bermakna atau tidak rancu.

## 1.7 Sistematika Penulisan

Untuk mengetahui uraian singkat yang memuat proses pembahasan dalam penelitian masing-masing bab yang dibuat, maka dibuat sistematika penulisan sebagai berikut:

## BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang pengantar terhadap dasar permasalahan yang akan dibahas dalam penelitian ini. Dasar permasalahan ini berisi latar belakang,

rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan dalam penulisan laporan ini.

## **BAB II TINJAUAN PUSTAKA**

Pada bab ini berisi kajian pustaka yang mirip dan pernah dilakukan sebelum penelitian ini. Bab ini juga menjelaskan teori yang digunakan untuk mendukung penelitian ini.

## **BAB III METODE PENELITIAN**

Pada bab ini membahas tentang alur penelitian yang berupa perancangan perangkat dan langkah penyelesaian sistematis dalam menyelesaikan permasalahan yang ada.

## **BAB IV HASIL DAN PEMBAHASAN**

Pada bab ini berisi tentang hasil dan pembahasan penelitian yang dilakukan secara lengkap termasuk hasil dari pengujian sistem.

## **BAB V PENUTUP**

Bab ini merupakan bagian terakhir dari penelitian yang berisi tentang kesimpulan dari seluruh bab yang sudah dibahas dan juga saran yang diuraikan untuk penelitian selanjutnya.