

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi dan informasi saat ini telah mengalami percepatan, terutama dengan pengaruh besar dari internet [1]. Fenomena ini tidak hanya membawa perubahan besar dalam cara manusia berinteraksi dan mengakses informasi, tetapi juga mengubah secara mendasar bagaimana bisnis, pendidikan, dan kehidupan sehari-hari dijalankan [2]. Salah satu hasil yang paling menonjol dari perkembangan teknologi ini adalah kemunculan media sosial. Twitter telah menjadi tempat utama bagi jutaan orang untuk berbagi pengalaman, ide, dan opini mereka terhadap berbagai macam permasalahan [1][3]. Twitter merupakan layanan jejaring sosial yang digunakan penggunanya untuk bertukar dan berbagi informasi mengenai banyak hal yang dapat menjadi perhatian. Data di media sosial Twitter yang berupa teks, maka data ini dapat di olah menjadi sebuah informasi dan pengetahuan yang berguna terhadap isu yang berkembang [1][3].

Indonesia adalah salah satu negara dengan pertumbuhan pengguna internet tercepat di dunia, dengan 170 juta dari 274,9 juta penduduknya atau sekitar 61,8% menjadi pengguna aktif media sosial. Dari jumlah tersebut, sekitar 63,6% menggunakan Twitter. [4]. Ini menunjukkan betapa pentingnya peran media sosial dalam kehidupan sehari-hari masyarakat Indonesia. Penggunaan Twitter tidak hanya sebatas untuk mengikuti teman-teman atau selebriti favorit, tetapi juga sebagai sarana untuk berdiskusi tentang isu-isu penting, mempromosikan produk atau acara, dan bahkan menyampaikan pandangan politik [1][3][4][5][6]. Namun, pertumbuhan pesat ini juga membawa tantangan baru, yaitu banyaknya informasi sering membuat pengguna kesulitan menemukan inti topik atau informasi yang relevan [LAMPIRAN A]

Banyaknya informasi yang tersedia sering kali menyulitkan masyarakat untuk menemukan inti informasi atau topik penting. Seperti saat mencari informasi di Twitter dengan kata kunci tertentu. Banyak tweet terkait kata kunci tersebut, dan

dalam satu tweet bisa mengandung banyak topik ditambah lagi dalam beberapa tweet dapat memiliki topik yang sama, ini membuat pengguna kesulitan mencari inti informasi. Ini menjadi tantangan dalam mengolah dan memahami data, bagi mereka yang memanfaatkan media sosial sebagai media informasi maupun untuk keperluan bisnis. Kondisi ini memiliki dampak yang signifikan, mulai dari mengacaukan pemahaman hingga mempengaruhi keputusan-keputusan penting salah satu contohnya, seorang *social media specialist* yang ingin mengetahui cara pemasaran yang sedang tren di media sosial jika *social media specialist* tersebut lambat dan sulit dalam memahami informasi yang ada di media sosial, ini akan berdampak buruk pada pengambilan keputusan di perusahaannya atau menyebabkan kerugian pada bisnisnya [6]. Dalam beberapa penelitian yang dilakukan oleh Swaminathan yang berjudul “*The Language of Brands in Social Media: Using Topic modeling on Social Media Conversations to Drive Brand Strategy*” telah dilakukan pendekatan *topic modeling* untuk mengatasi masalah ini dan terbukti hasilnya cukup baik, namun ada beberapa kekurangan, yaitu topik yang dihasilkan hanya berupa larik kata yang diinterpretasi secara manual oleh peneliti sebelumnya, yang menjadikan hasil ini sulit dipahami oleh pengguna, selain itu penelitian sebelumnya mengambil kasus yang spesifik dalam sebuah bidang seperti politik, bisnis, dan lain-lain [5][6].

Untuk mengatasi tantangan ini, teknik-teknik penambangan data atau *data mining* seperti *Aggregated Topic Models* dengan menggunakan algoritma *Latent dirichlet allocation* (LDA) serta menginterpretasi dengan *Large language model* (LLM) menjadi solusi yang menarik. *Data mining* adalah proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting pada data [7]. Penambangan data dalam bentuk teks biasanya dikenal juga dengan *Text Mining* [8], Algoritma LDA termasuk dalam *unsupervised learning* atau lebih tepatnya algoritma ini termasuk dalam *soft clustering*, algoritma ini, dapat melakukan pengelompokan topik utama dalam kumpulan teks. Dengan menerapkan algoritma LDA pada data tweet dan melakukan interpretasi dengan LLM, diharapkan masyarakat dapat dengan lebih mudah memahami topik-topik utama yang sedang dibicarakan di Twitter. Penelitian ini diharapkan dapat

memberikan kontribusi yang signifikan dalam bidang *data mining*, *text mining* dan analisis media sosial, serta membantu masyarakat secara individu atau kelompok dalam menghadapi tantangan dalam memahami informasi di era digital ini.

1.2 Perumusan Masalah

Berdasarkan penjelasan pada latar belakang masalah sebelumnya, perumusan masalah pada penelitian ini apakah dengan menerapkan *Aggregated Topic Models* dari data media sosial X (Twitter) dapat menghasilkan topik-topik yang sesuai dengan kebutuhan pengguna?

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk menerapkan *Aggregated Topic Models* dan Interpretasi LLM untuk mengetahui topik dari media sosial X (Twitter) berdasarkan *keyword* secara *general*. Tujuan dari penelitian ini adalah membantu pengguna untuk mendapatkan topik yang mudah dipahami berdasarkan data sosial media X (Twitter) dalam rentang waktu tertentu.

1.4 Batasan Masalah

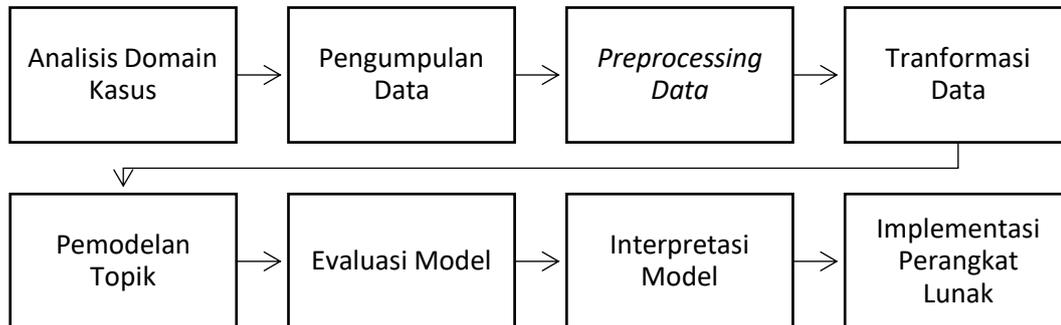
Dalam melakukan penelitian ada beberapa batasan masalah menjadi yang diterapkan sebagai berikut:

- A. Data yang digunakan adalah data tweet dari hasil *crawling* Twitter dengan Tweet Harvest
- B. Pengumpulan data pada Twitter hanya dapat data tweet rentang 7 hari sebelumnya dengan batas jumlah tweet 10.000.
- C. Tweet yang diproses hanya tweet yang berbahasa Indonesia.
- D. Sistem yang dibuat berbentuk web

1.5 Metodologi Penelitian

Penelitian ini akan mengadaptasi metodologi *Knowledge Discovery in Database* (KDD) [9], yang telah disesuaikan dengan kebutuhan dan implementasi.

Adapun hasil adaptasi metodologi ini, dijelaskan dalam diagram pada Gambar 1.1 Metodologi Penelitian



Gambar 1.1 Metodologi Penelitian

Berikut merupakan penjelasan dari langkah-langkah yang terdapat pada penelitian ini:

1. Analisis Domain Kasus

Analisis Domain Kasus adalah proses menentukan ruang lingkup dari topik yang dibutuhkan masyarakat secara individu dalam domain kasus yang lebih *general* atau umum. Dalam menganalisis, permasalahan yang berkembang di masyarakat terhadap beberapa bidang tertentu.

2. Pengumpulan Data

Pengumpulan data dilakukan pada media sosial Twitter dengan menggunakan Twitter Harvest, alat yang dirancang untuk mengambil data dari Twitter secara otomatis. Parameter pencarian disesuaikan berdasarkan kata kunci seperti politik, kesehatan, ekonomi dan pendidikan dengan kebutuhan dari rentang waktu hingga penyesuaian bahasa, mencakup isi tweet, *hashtag*, atau akun-akun pengguna tertentu yang relevan dengan penelitian, karena domain yang diangkat adalah hal umum.

3. Data Preprocessing

Preprocessing adalah tahapan dalam analisis data yang bertujuan untuk mempersiapkan data mentah agar dapat diolah lebih lanjut oleh algoritma atau model analisis data. Data yang diambil dari tahap sebelumnya akan menjalani proses *Preprocessing* yang mencakup *case folding*, *cleansing*, *tokenizing*, *Normalization*, dan *stopword removal*, yang disesuaikan dengan persyaratan analisis *Topic modeling*.

4. Transformasi Data

Data yang telah dilakukan *preprocessing* pada tahap sebelumnya akan dilakukan transformasi atau ekstraksi fitur dalam kasus ini data hasil *preprocessing* akan dilakukan perhitungan *Bag of Word*. Pada tahap ini akan diidentifikasi kata unik yang ada dalam sebuah dokumen, kemudian berdasarkan kata unik tersebut akan ditentukan kemunculannya setiap di dalam sebuah dokumen. Data ini nantinya akan dibuat matriks dokumen terhadap topik dan topik terhadap kata.

5. Pemodelan Topik (*Topic Modeling*)

Data yang telah dilakukan transformasi pada tahap sebelumnya akan dilakukan pemodelan topik yaitu mengidentifikasi topik-topik dominan dalam data tersebut dan mengelompokkan data tersebut ke dalam kategori topik, dalam penelitian ini pemodelan topik akan dilakukan dengan menggunakan *Aggregated Topic Models* dengan algoritma *Latent dirichlet allocation* (LDA).

6. Interpretasi Hasil Topik

Interpretasi hasil topik adalah proses menganalisis dan memahami makna dari topik-topik yang dihasilkan dari model pemodelan topik. Dalam menginterpretasi topik dengan bentuk larik kata akan dilakukan dengan bantuan LLM untuk dibuat sebuah kalimat yang padu. interpretasi hasil topik memainkan peran penting dalam menyampaikan wawasan yang berharga dari data teks dan membantu pengguna memahami konteks dan implikasi dari temuan yang dihasilkan.

7. Evaluasi Model

Hasil dari pemodelan topik akan dievaluasi dengan cara melihat apakah topik yang dihasilkan sudah sesuai atau memiliki nilai koherensi yang baik. Dengan evaluasi koherensi ini dapat melihat berapa nilai keterhubungan kata-kata dengan topik serta data yang ada nilai koherensi memiliki rentang 0-1 semakin tinggi nilai koherensi menandakan topik semakin baik. Topik yang telah dihasilkan dari proses LDA pada tahap sebelumnya akan dicek kembali dengan evaluasi koherensi untuk memastikan topik tersebut sudah baik atau belum.

8. Implementasi Perangkat Lunak

Implementasi perangkat lunak adalah proses mengubah desain atau konsep perangkat lunak menjadi bentuk yang dapat dioperasikan atau digunakan secara

nyata. Dalam tahap ini akan diimplementasi model algoritma dan metode yang digunakan ditahap sebelumnya menjadi Rest-API yang digunakan pada aplikasi *website*.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran umum mengenai penelitian yang dikerjakan. Sistematika penulisan dalam tugas akhir ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Pada bab 1 membahas tentang latar belakang, Perumusan masalah, menjelaskan maksud dan tujuan, menguraikan batasan masalah, menjelaskan metodologi penelitian yang digunakan, dan menjelaskan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Pada bab 2 membahas tentang uraian hasil yang didapatkan pada penelitian yang dilakukan, konsep dasar, dan teori dari para ahli yang berkaitan dengan penelitian. Meninjau permasalahan dan hal-hal yang berguna dari penelitian-penelitian serupa yang pernah dikerjakan sebelumnya, lalu menggunakannya untuk acuan dalam pemecahan masalah pada penelitian ini.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Pada bab 3 membahas tentang analisis domain kasus, perancangan, serta evaluasi model hasil dari penelitian untuk mengetahui masalah apa yang muncul lalu mencoba untuk memecahkan masalah tersebut dengan *Aggregated Topic Models* dan LLM.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada bab 4 membahas tentang implementasi dari masalah yang telah dianalisis sebelumnya, selain itu dilakukan penanganan dengan menggunakan yang metode telah ditentukan, dan dilakukan pengujian pada hasil penanganannya.

BAB 5 KESIMPULAN DAN SARAN

Pada bab 5 membahas tentang kesimpulan yang didapatkan dari hasil penelitian ini dan saran terhadap pengembangan selanjutnya.