

# BAB I PENDAHULUAN

## 1.1.Latar Belakang

Di era digital yang semakin berkembang, survei *online* telah menjadi salah satu metode yang paling efisien dan populer untuk mengumpulkan data dalam berbagai bidang, termasuk pendidikan, bisnis, dan pemerintahan [1]. Survei *online* memberikan keuntungan, seperti fleksibilitas, biaya rendah, serta kemampuan untuk menjangkau populasi yang luas dan memperoleh data dengan jumlah besar dalam waktu singkat. Responden dapat mengisi survei kapan saja dan di mana saja, tanpa batasan geografis, sehingga memungkinkan peneliti untuk mendapatkan data yang lebih representatif dari populasi yang lebih luas. Melalui platform survei *online* seperti *Google Forms*, *SurveyMonkey*, atau platform institusi, penyelenggara dapat dengan mudah mendistribusikan survei kepada peserta sebelum, selama, atau setelah acara berlangsung.

Top In Feedback Forms and Surveys Usage Distribution in Indonesia		
Technology	Websites	%
Contact Form 7	50,044	55.71
WPForms	20,663	23
MailChimp	3,360	3.74
Hotjar	2,603	2.9
Trustpilot	1,556	1.73
TrustIndex	1,440	1.6
OptinMonster	1,423	1.58
Google Forms	984	1.1
CodeMirror	864	0.96
Zendesk Embeddables	803	0.89

**Gambar 1.1** Formulir *Feedback* dan Distribusi Penggunaan Survei Terpopuler di Indonesia [2]

Berdasarkan data statistik tahun 2024 dari situs web Gambar 1.1, Google Forms termasuk dalam 8 besar platform formulir feedback dan survei terpopuler di Indonesia, dengan 984 situs web yang tercatat menggunakan layanan ini. Popularitas Google Forms yang cukup tinggi di Indonesia menunjukkan keunggulannya dalam menyediakan solusi praktis untuk pembuatan survei online dan pengumpulan data, baik untuk keperluan bisnis, akademik, maupun organisasi

lainnya. Hal tersebut menjadikan pilihan yang menarik dan relevan sebagai objek penelitian bagi penulis.

Google Forms merupakan salah satu instrumen survei online untuk mengumpulkan feedback, baik dalam bentuk pertanyaan tertutup (*close-ended*) dan terbuka (*open-ended*). Pertanyaan *close-ended* menyediakan pilihan jawaban yang sudah ditentukan, sehingga memudahkan proses analisis karena datanya terstruktur dan dapat divisualisasikan dengan cepat dalam bentuk grafik atau tabel [3]. Pertanyaan *close-ended* menghasilkan data yang lebih cepat diproses dibandingkan dengan pertanyaan *open-ended*, khususnya dalam survei skala besar [4]. Sebaliknya, pertanyaan *open-ended* memberikan kebebasan bagi responden untuk menjawab dengan kata-kata mereka sendiri, memungkinkan pengumpulan informasi yang lebih mendalam [5]. Data yang dihasilkan dari pertanyaan *open-ended* tidak terstruktur dan harus dianalisis untuk mengekstrak informasi yang bermanfaat. Sebagai contoh, Gambar 1.2 menunjukkan jawaban dari pertanyaan *open-ended* yang diperoleh dari feedback seminar nasional FKBM-IK tahun 2022 yang berisi kritik dan saran dari peserta. Hal ini menjadi tantangan bagi penyelenggara kegiatan, terutama ketika jumlah responden sangat besar.

G	H	I	J	K	L	M	N
Kritik dan saran :							
-							
.							
buat webinar lagi							
jaringan lebih di perbaiki, meskipun sedikit nge-lag sedikit tadi waktu acara dimulai							
Tetap menyelenggarakan event FKBM-IK ditahun-tahun selanjutnya dengan membawa materi yang edukatif :)							
Tolong anggap saya							
-							
semogaa bisa ga telat waktunya							
Sudah keren dan menambah insight. Mungkin ke depannya, dapat dipertahankan kualitas webinarnya.							
tidak ada							
Sudah cukup baik.							
lebih banyak lagi diadain webinar yang sangat bermanfaat seperti ini dan mengundang speaker yang tidak kalah keren kedepannya							
-							
Untuk webinar pagi ini berjalan dengan baik namun untuk memulai acaranya dapat tingkatkan lagi untuk ketepatan waktunya.							
Tidak ada.							
gadaa, seruuu pol							
kerenn!!							

**Gambar 1.2 Form *Feedback Open-Ended* (FKBM-IK, 2022)**

Google forms tidak menyediakan fitur yang memadai untuk pengelolaan data *open-ended* dengan efisien. Data *open-ended* sering kali berisi informasi penting yang bisa diperoleh dari respons pengguna. Namun, keterbatasan fitur untuk pengelolaan data ini membuat analisis manual menjadi tidak efektif, terutama

ketika jumlah responden sangat besar, karena analisis manual membutuhkan waktu dan tenaga yang besar untuk membaca dan mengklasifikasikan sentimen. Selain itu, karena data yang diperoleh tidak terstruktur, diperlukan pendekatan pemrosesan yang lebih kompleks seperti *Natural Language Processing* (NLP) atau visualisasi data. Pendekatan ini dapat membantu untuk mempercepat dan menyederhanakan analisis data open-ended, terutama dalam survei skala besar, sehingga penyelenggara dapat memperoleh wawasan yang bermanfaat dari hasil survei tersebut [6].

Salah satu bidang dalam NLP yang dapat mengelola data tidak terstruktur adalah analisis sentimen. Analisis sentimen atau penambangan opini merupakan salah satu bidang yang berkembang dalam pemrosesan bahasa alami [7]. [8] mengatakan bahwa tugas utama analisis sentimen yaitu mengklasifikasikan pendapat yang terdapat dalam teks. [9] menjelaskan bahwa opini yang diekstraksi biasanya diklasifikasikan berdasarkan polaritasnya, yaitu positif, negatif, atau netral. Menurut [10] analisis sentimen memiliki tiga tingkat klasifikasi, yaitu dokumen, kalimat, dan aspek.

Analisis sentimen sering digunakan dalam pengelolaan data tekstual untuk memahami persepsi pengguna. Meskipun teknik ini dapat memberikan gambaran umum mengenai sentimen positif, negatif, atau netral, pendekatan ini sering kali kurang informatif ketika diterapkan pada data *open-ended*, seperti survei atau ulasan yang mencakup banyak aspek dan dimensi. Misalnya, ulasan tentang sebuah acara konferensi, pengguna mungkin menyukai kontennya tetapi merasa tidak puas dengan pembicaranya. Dalam hal ini, analisis sentimen tradisional hanya akan memberikan gambaran keseluruhan (positif, negatif, atau netral), tanpa mengidentifikasi sentimen pada aspek-aspek spesifik seperti konten, fasilitas, atau pembicara [7].

Dalam mengatasi keterbatasan analisis sentimen, *aspect-based sentiment analysis* (ABSA) menjadi solusi yang lebih efektif untuk memberikan informasi sentimen secara mendetail. Menurut [11] pendekatan ini sangat bermanfaat di berbagai domain, seperti *e-commerce*, perhotelan, dan pendidikan untuk menganalisis opini dari ulasan pengguna terhadap produk atau layanan [12]. ABSA dapat menguraikan respons ke dalam berbagai aspek spesifik, seperti konten,

fasilitas, atau pembicara. Setiap aspek dapat dianalisis sentimennya secara terpisah. Menurut [13], ABSA memberikan pandangan yang lebih dalam terhadap opini pengguna karena mampu memisahkan sentimen pada level aspek, sehingga menghasilkan analisis yang lebih kaya dibandingkan pendekatan sentimen tradisional yang hanya membedakan antara positif, negatif, dan netral secara keseluruhan. Penelitian yang dilakukan oleh [14] menunjukkan bahwa ABSA efektif dalam mengidentifikasi sentimen berdasarkan aspek tertentu dalam data tidak terstruktur. Studi ini memberikan hasil yang signifikan dalam meningkatkan akurasi deteksi aspek dan sentimen di berbagai domain, seperti ulasan restoran dan produk.

Mengidentifikasi aspek pada teks tidak terstruktur seperti respons survei open-ended bukanlah tugas yang mudah. *Latent Dirichlet Allocation (LDA)*, sebagai salah satu metode pemodelan topik yang populer, sering digunakan untuk mengklasifikasikan teks menjadi beberapa topik atau aspek utama. Menurut [15] dalam [16], LDA dapat mengidentifikasi pola distribusi kata dalam dokumen dan mengklasifikasikan ke dalam topik tertentu. Setelah aspek dalam dokumen teridentifikasi, tahap selanjutnya adalah menganalisis sentimen yang terkait dengan aspek tersebut.

Dalam ABSA, pemahaman konteks kata sangat penting agar model dapat menginterpretasikan opini secara akurat. Untuk itu, diperlukan teknik representasi teks yang kaya dan kontekstual. Salah satu teknik yang banyak digunakan untuk representasi teks adalah *word embedding*, yang mengubah kata-kata menjadi vektor numerik berdasarkan hubungan semantik dan sintaksis dalam korpus pelatihan. *Bidirectional Encoder Representations from Transformers (BERT)* merupakan salah satu pendekatan berbasis *deep learning* yang terbukti telah terbukti memberikan hasil yang lebih baik dalam memahami makna kata berdasarkan konteksnya dibandingkan metode tradisional seperti Word2Vec atau GloVe [17].

Pada penelitian ini akan melibatkan dua tahapan. Pertama, *clustering topic* atau aspek akan dilakukan menggunakan metode LDA seperti penelitian [18] yang menganalisis sentimen ulasan hotel untuk menentukan dan memahami jumlah topik dalam dataset. Kedua, klasifikasi sentimen berdasarkan aspek akan dilakukan menggunakan *Support Vector Machine (SVM)* karena SVM telah terbukti lebih

unggul dengan pendekatan yang lebih matang, terdefinisi secara matematis, serta memiliki dasar teori dan kapasitas generalisasi yang baik dibandingkan dengan metode klasifikasi lainnya [19]. Melalui tahapan ini, diharapkan respons survei *open-ended* dapat dianalisis dengan lebih akurat dan mendalam, sehingga menghasilkan wawasan yang bermanfaat terkait persepsi dan opini responden terhadap berbagai aspek acara atau kegiatan yang dievaluasi, seperti kualitas penyelenggaraan, konten acara, kepuasan terhadap narasumber, dan pengalaman secara keseluruhan.

### **1.2.Rumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan di atas, diperoleh rumusan masalah sebagai berikut:

1. Bagaimana hasil penerapan LDA, BERT *Embedding*, dan SVM untuk menangani data survei *open-ended*?
2. Bagaimana konfigurasi model yang memiliki performa paling optimal untuk menangani data survei *open-ended*?

### **1.3.Batasan Masalah**

Batasan masalah digunakan untuk menjaga agar penelitian tetap fokus dan terstruktur sesuai tujuan penulis. Adapun batasan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya berfokus pada respons survei *open-ended* berbahasa Indonesia.
2. Penelitian ini menggunakan respons survei *open-ended* dari kegiatan seminar nasional FKBM-IK tahun 2021 dan 2022.
3. Model BERT yang digunakan adalah IndoBERT (*indobert-base-p2*).
4. Analisis sentimen yang dilakukan terbatas pada beberapa aspek dominan dari survei *open-ended*.

#### **1.4. Tujuan Penelitian**

Tujuan dari penelitian ini yaitu:

1. Mengembangkan sistem ABSA untuk menganalisis respons survei open-ended menggunakan pendekatan LDA, BERT *Embedding* dan SVM.
2. Mengetahui efektivitas model yang dikembangkan dalam mengklasifikasikan sentimen berbasis aspek pada respons survei *open-ended*.

#### **1.5. Manfaat Penelitian**

Melalui penelitian ini diharapkan bermanfaat untuk perusahaan atau organisasi dalam menganalisis teks dan mengidentifikasi sentimen dari berbagai elemen (*aspect*). Selain itu, penelitian ini juga dapat membantu dalam memahami topik atau aspek dari respons survei dengan lebih efektif, mendukung proses pengambilan keputusan, serta mengidentifikasi kelebihan dan kekurangan agar dapat meningkatkan kualitas atau layanan.