ANALISIS SENTIMEN BERBASIS ASPEK PADA RESPONS SURVEI *OPEN-ENDED* MENGGUNAKAN LDA, DAN SVM

Dian Rahmawati, Eka Dyar Wahyuni, Dhian Satria Yuda Kartika
Sistem Informasi, UPN "Veteran" Jawa Timur
Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Indonesia
diianrhma@gmail.com

ABSTRAK

Dalam sebuah acara seminar nasional, evaluasi kepuasan peserta menjadi aspek penting untuk menilai keberhasilan acara dan mengusulkan perbaikan di masa mendatang. Salah satu metode evaluasi yang umum digunakan adalah survei *online*, yang sering kali mencakup pertanyaan terbuka (*open-ended*) agar peserta dapat menyampaikan opini dan pengalaman mereka secara bebas. Namun, respons survei *open-ended* bersifat tidak terstruktur, sehingga analisis manual menjadi kurang efektif karena memerlukan banyak waktu dan rentan terhadap subjektivitas. Analisis mendalam terhadap respons survei *open-ended* diperlukan untuk memahami opini peserta mengenai kualitas acara. Penelitian ini bertujuan untuk mengklasifikasikan sentimen dan mengidentifikasi aspek utama dalam respons survei *open-ended* dari peserta seminar. Penelitian ini menggunakan Metode *Latent Dirichlet Allocation* (LDA) untuk *topic modeling*, dan *Support Vector Machine* (SVM) untuk klasifikasi sentimen. Hasil penelitian menunjukkan bahwa kombinasi SVM dan teknik *resampling Synthetic Minority Over-sampling Technique* (SMOTE) menghasilkan performa terbaik dengan akurasi 85%, Precision 0.78, Recall 0.88, dan F1-Score 0.81. Analisis LDA berhasil mengidentifikasi 9 topik utama yang mencerminkan aspek penting dalam respons peserta, dengan mayoritas sentimen yang ditemukan bersifat positif, terutama dalam topik Manfaat atau Materi.

Kata kunci : *Analisis Sentimen, LDA, SVM, Survei Open-ended*

1. PENDAHULUAN

Di era digital, survei online menjadi metode efektif untuk mengumpulkan data di berbagai sektor karena fleksibilitas, jangkauan luas, dan efisiensi biaya [1].

Berdasarkan data statistik 2024 [2], Google Forms merupakan salah satu platform survei terpopuler di Indonesia, yang digunakan oleh 984 situs web seperti yang terlihat dalam Gambar 1.

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. Data Statistik 2024 Penggunaan Survei di Indonesia

Google Forms menyediakan dua jenis pertanyaan, yaitu tertutup (*close-ended*) dan terbuka (*open-ended*) [3].

Pertanyaan *close-ended* lebih mudah dianalisis karena datanya terstruktur. Sebaliknya, pertanyaan *open-ended* memberikan wawasan lebih mendalam, tetapi sulit diproses karena tidak terstruktur dan memerlukan analisis manual yang memakan waktu [4].

Google Forms belum memiliki fitur yang memadai untuk menganalisis data *open-ended*, sehingga diperlukan metode seperti *Natural Language Processing* (NLP) guna meningkatkan efisiensi pengelolaan data dalam skala besar [5].

Salah satu teknik NLP yang umum digunakan adalah analisis sentimen, yang mengklasifikasikan opini menjadi positif, negatif, atau netral [6].

Namun, metode ini kurang efektif untuk survei *open-ended* karena tidak dapat mengidentifikasi sentimen pada aspek spesifik. *Aspect-Based Sentiment Analysis* (ABSA) menjadi solusi dengan memisahkan sentimen berdasarkan aspek tertentu [7].

Untuk mengidentifikasi aspek dalam teks tidak terstruktur, dapat digunakan *Latent Dirichlet Allocation* (LDA), yang mengelompokkan kata dalam topik tertentu [8].

Setelah aspek teridentifikasi, kemudian, *Support Vector Machine* (SVM) digunakan untuk klasifikasi sentimen karena keunggulannya dalam generalisasi dan pendekatan matematis [9] dalam [10].

2. TINJAUAN PUSTAKA

Pada tahap ini, akan disajikan beberapa literatur ilmiah yang berkaitan dengan penelitian.

2.1. Penelitian Terdahulu

Penelitian [11] menganalisis sentimen berbasis aspek pada ulasan aplikasi EdLink di Google Play Store menggunakan *Latent Dirichlet Allocation* (LDA) dan *Support Vector Machine* (SVM) dengan pendekatan Lexicon Based. Data yang digunakan mencakup 2014 ulasan dari versi 1.1.6 hingga 4.7.8.

Hasil pemodelan LDA mengidentifikasi tiga aspek utama: *Application Usability*, *Reliability*, dan *Performance Efficiency*, dengan skor koherensi tertinggi 0,487. *Labeling Lexicon Based* menghasilkan 418 ulasan positif dan 1.223 ulasan negatif. Klasifikasi SVM dengan rasio data 90:10 mencapai akurasi 85,45%, meningkat menjadi 90,00% setelah resampling menggunakan SMOTE. Distribusi sentimen menunjukkan dominasi ulasan negatif di semua aspek, terutama pada *Reliability* (482 ulasan negatif) dan *Performance Efficiency* (422 ulasan negatif).

Selanjutnya, penelitian [12] menganalisis sentimen ulasan pengguna aplikasi Starbucks *Loyalty Rewards App* serta aspek usability seperti learnability, efficiency, errors, dan satisfaction. Data diperoleh melalui scraping ulasan di Google Play Store, kemudian diproses menggunakan TF-IDF untuk ekstraksi fitur. Metode Support Vector Machine (SVM) diterapkan dengan tiga jenis kernel (Linear, Polinomial, dan RBF), serta dilakukan Hyperparameter tuning menggunakan GridSearchCV untuk meningkatkan performa model. Hasil penelitian menunjukkan bahwa model SVM memiliki akurasi rata-rata 88,96%, f1-score 66,85%, precision 75,77%, dan recall 64,68%. Analisis sentimen mengungkap bahwa mayoritas ulasan bernilai negatif, terutama pada aspek errors, yang menunjukkan tingginya tingkat kesalahan sistem dalam aplikasi.

2.2. Survei Open-Ended

Survei merupakan metode yang paling efektif untuk memperoleh data dari sejumlah besar responden yang tersebar secara geografis [13].

Berdasarkan formatnya, pertanyaan survei dapat diklasifikasikan menjadi pertanyaan tertutup (*close-ended*) dan terbuka (*open-ended*). Survei *open-ended* merupakan jenis survei yang memungkinkan responden untuk memberikan jawaban bebas tanpa dibatasi oleh opsi jawaban yang telah ditentukan sebelumnya.

Pertanyaan *open-ended* memungkinkan pengumpulan data yang lebih mendalam karena responden dapat memberikan pandangan mereka secara bebas dan detail. Jawaban yang tidak terbatas pada pilihan yang telah ditentukan memungkinkan peneliti untuk menemukan perspektif baru atau isu yang sebelumnya tidak teridentifikasi.

2.3. Aspect-Based Sentiment Analysis (ABSA)

Analisis sentimen merupakan salah satu bidang penelitian dalam text mining yang bertujuan untuk mengidentifikasi aspek-aspek dari suatu entitas dan menentukan polaritas sentimen terhadap setiap aspek tersebut. Misalnya, ketika melihat ulasan tentang sebuah hotel, opini yang diberikan oleh masyarakat tidak hanya mencakup sentimen umum, tetapi juga mencakup berbagai aspek seperti lokasi, pelayanan, makanan, kenyamanan, dan lainnya [9].

ABSA mencakup dua tugas utama: (1) aspect extraction, yaitu mengekstraksi aspek yang relevan dari teks, dan (2) sentiment classification, yaitu menentukan polaritas sentimen (positif, negatif, atau netral) terhadap aspek yang telah diidentifikasi.

2.4. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) merupakan salah satu metode pemodelan topik yang banyak digunakan [14].

Pemodelan topik adalah teknik untuk klasifikasi dokumen tanpa pengawasan, mirip dengan pengelompokan data numerik, yang dapat mengidentifikasi beberapa kelompok topik meskipun tidak memiliki gambaran yang jelas tentang apa yang dicari.

Pemodelan topik bertujuan untuk menentukan aspek-aspek yang terdapat dalam ulasan. Salah satu faktor utama dalam pemodelan ini adalah jumlah topik, yang akan dievaluasi menggunakan topic coherence. Metrik ini mengukur tingkat kesamaan antara kata-kata dalam setiap topik, di mana nilai *topic coherence* yang lebih tinggi menunjukkan hasil yang lebih baik.

2.5. Confusion Matrix

Confusion matrix merupakan tabel yang merangkum kinerja model *machine learning* pada data uji. Ini merupakan metode sistematis untuk mengalokasikan prediksi ke kelas asli dari data yang awalnya ada. Matriks ini menunjukkan jumlah *true positives* (TP), *true negatives* (TN), *false positives* (FP), dan *false negatives* (FN) yang dihasilkan oleh model pada data uji. *Confusion matrix* digunakan untuk menghitung berbagai metrik kinerja model *machine learning*, seperti akurasi, presisi, recall, dan F1 score [15].

Tabel 1. merupakan contoh *confusion matrix* untuk klasifikasi biner:

Tabel 1. Confusion Matrix

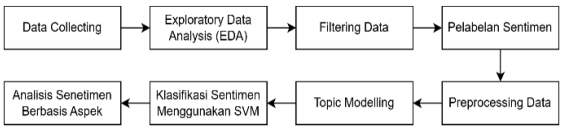
	Actual Positive	**Actual Negative**
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (FN)

2.6. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu metode *supervised learning* yang sering digunakan untuk klasifikasi dan regresi. Dalam pemodelan klasifikasi, SVM telah terbukti lebih unggul karena memiliki pendekatan yang lebih matang dan lebih terdefinisi secara matematis dibandingkan dengan metode klasifikasi lainnya. SVM juga sering digunakan sebagai metode klasifikasi karena dinilai memiliki dasar teori dan kapasitas generalisasi yang baik [10].

3. METODE PENELITIAN

Pada bab metodologi penelitian, menjelaskan tahapan yang digunakan dalam penelitian agar gar tersusun secara terstruktur dan sistematis. Gambar 2 menunjukkan alur penelitian yang dilakukan.



Gambar 2. Alur Penelitian

3.1. Data Collecting

Data Collecting dilakukan melalui penyebaran survei *online* menggunakan platform Google Forms pada acara seminar nasional FKBM-IK selama dua periode yaitu tahun 2021 dan 2022. Penulis hanya menggunakan satu kolom, yaitu kritik dan saran.

3.2. Exploratory Data Analysis (EDA)

Pada tahap EDA, digunakan untuk memahami karakteristik, mengidentifikasi pola, mengeksplorasi hubungan antar variabel, serta mendeteksi potensi masalah dalam data sebelum melakukan model prediktif atau analisis lebih lanjut dalam bentuk *word cloud*.

3.3. Penyaringan Data

Penyaringan data bertujuan untuk membersihkan dataset dari entri yang pendek serta tidak relevan atau tidak informatif seperti 'tidak ada', 'terima kasih', dsb. Dalam konteks ini, entri yang mengandung kurang dari dari tiga karakter serta tidak memberikan informasi yang cukup akan dihapus dari dataset dan tidak diproses lebih lanjut.

3.4. Pelabelan Sentimen

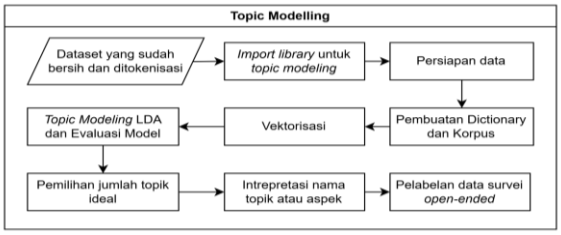
Pelabelan sentimen bertujuan untuk mengidentifikasi sentimen dari respons survei *open-ended*. Tahapan pelabelan sentimen diawali dengan translasi teks survei *open-ended* dari bahasa Indonesia ke bahasa Inggris. Translasi dilakukan karena keempat pelabel otomatis yang digunakan Vader, TextBlob, CardifF, DistilBERT, dan palabel manual.

3.5. Data Preprocessing

Tahap data *preprocessing* bertujuan untuk menghilangkan *noise*, karena data harus dalam kondisi bersih dan terstruktur sebelum dilakukan pemodelan. Tahapan *data preprocessing* meliputi, *data cleaning*, *case folding*, *tokenization*, *normalization*, dan *stopword removal*.

3.6. Topic Modeling

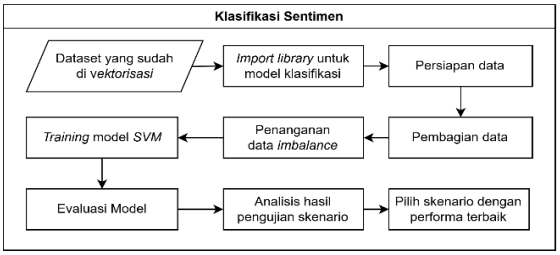
Topic modeling dengan model *Latent Dirichlet Allocation* (LDA) digunakan untuk mengidentifikasi topik dalam data respons survei *open-ended*. Model ini bertujuan untuk memetakan distribusi topik dalam dokumen dengan penyesuain parameter yang diperlukan untuk memperoleh jumlah topik yang optimal atau ideal. Gambar 3 menunjukkan alur dari topic modeling menggunakan LDA.



Gambar 3. Alur Topic Modeling

3.7. Klasifikasi Sentimen *Support Vector Machine*

Tahap klasifikasi sentimen terdiri dari beberapa tahapan, seperti pembagian data dengan *holdout*, klasifikasi sentimen menggunakan SVM, evaluasi dan evaluasi klasifikasi dengan *confusion matrix*. Gambar 4. Menunjukan alur dari klasifikasi sentimen menggunakan SVM.



Gambar 4. Alur Klasifikasi Sentimen

3.8. Analisis Sentimen Berbasis Aspek (ABSA)

Setetlah proses topic modeling LDA dan klasifikasi sentimen menggunakan SVM selesai, dilakukan analisis hasil untuk mengetahui distribusi setiap aspek dan sentimen. Hasil analisis tersebut dapat digunakan sebagai bahan evaluasi untuk perbaikan atau peningkatan kualitas acara di masa mendatang.

4. HASIL DAN PEMBAHASAN

Bab ini, memuat, hasil, pengujian, serta pembahasan yang akan dijelaskan lebih lanjut pada bagian berikut.

4.1. Data Collecting

Data yang digunakan dalam penelitian ini merupakan feedback seminar nasional FKBM-IK 2021 dan 2022. Data ini diperoleh menggunakan alat survei *online* Google Forms dengan mencakup beberap variabel, namun data yang digunakan dalam analisis hanya satu kolom kritik dan saran.

1	feedback
2	jaringan lebih di perbaiki, meskipun sedikit nge-lag sedikit tadi waktu acara dimulai
3	Tetap menyelenggarakan event FKBM-IK ditahun-tahun selanjutnya dengan membawa materi yang edukatif :)
4	sudah bagus,kurang games sAJA
5	Materi yang dibawakan cukup menarik. Untuk panitia sebaiknya acara dimulai dan di akhiri tepat waktu. Terimakasih
6	semoga lebih ontime lagi
7	Kegiatan FKBMIK berjalan dengan baik, narasumber membawa materi yang seru dan sangat relatable dengan masa-masa sekarang.
8	Kapasitas zoomnya ditambah kalo bisa
9	Saran saja mungkin untuk waktu pelaksanaan lebih bisa tepat waktu agar tidak molor
10	Interaksi terhadap pendengar lebih banyak lagi
11	mungkin diadakan games agar lebih seru tidak monoton
12	

Gambar 5. Dataset Penelitian

Gambar 3 menunjukkan dataset yang digunakan dalam penelitian, dengan total 1.276 responden. Data tersebut kemudian disimpan dalam format CSV untuk memudahkan analisis lebih lanjut.

4.2. Exploratory Data Analysis (EDA)

Selanjutnya, proses visualisasi menggunakan word cloud untuk memahami karakterisitik, pola, serta mendeteksi potensi masalah yang terdapat dalam dataset.

Gambar 6. Word Cloud Dataset Awal

Berdasarkan hasil *word cloud* Gambar 4, dapat dilihat bahwa masih banyak kata-kata yang tidak relevan untuk analisis seperti “dan”, “yang”, “ini”, dsb. Sehingga, perlu penanganan lebih lanjut pada tahap berikutnya.

4.3. Penyaringan Data

Tahap penyaringan data dilakukan secara manual, dengan membaca setiap baris respons survei *open-ended.* Contoh data yang dilakukan penyaringan adalah data yang mengandung emoji/emoticon, entri pendek kurang dari tiga karakter, dsb.

Tabel 2. Data yang dilakukan Penyaringan

Sebelum	Sesudah
Tidak ada	(dihapus)
terimakasih	(dihapus)
👍👍👍👍👍	(dihapus)

Tabel 2. Merupakan data yang dilakukan penyaringan, sehingga jumlah data yang tersisa sebanyak 697 data dari 1.276 data awal.

4.4. Pelabelan Sentimen

Pelabelan sentimen dilakukan oleh lima pelabel yaitu empat pelabel otomatis (Vader, TextBlob, Cardiff, DistilBert), dan satu pelabel manual. Kelas yang digunakan dalam pelabelan sentimen adalah positif, atau negatif. Tabel 3. Merupakan hasil dari pelabelan sentimen.

Tabel 3. Hasil Pelabelan Sentimen

Kelas	Jumlah
Positif	547
Negatif	110
Total	697

.

4.5. Data Preprocessing

Tahap ini memanfaatkan library Python serta mencakup beberapa tahapan seperti *data cleaning, case folding, tokenization, normalization,* dan *stopword removaL.*

4.5.1. Data Cleaning

Tahap ini bertujuan untuk menghapus emoji atau emoticon, *mention*, *link*, hashtag, angka, tanda baca, spasi di awal atau di akhir (*whitespace*), serta karakter yang berulang. Tabel 4. Merupakan hasil dari *data cleaning*.

Tabel 4. Hasil Data Cleaning

Sebelum	Sesudah
acaranya sudah bagus dan lebih dikembangkan lagi #respect	acaranya sudah bagus dan lebih dikembangkan lagi

4.5.2. Case Folding

Tahap ini bertujuan untuk mengubah seleuruh format teks menjadi huruf kecil (*lowercase*) agar seragam. Tabel 5. Merupakan hasil dari case folding.

Tabel 5. Hasil Case Folding

Sebelum	Sesudah
kurang games sAJA	kurang games saja

4.5.3. Tokenization

Tahap ini bertujuan untuk memecah teks menjadi token atau kata-kata individual untuk memudahkan analisis selanjutnya. Tabel 6. Merupakan hasil dari tokenization.

Tabel 6. Hasil Tokenization

Sebelum	Sesudah
sudah bagus pembicara dan narasumbernya	['sudah', 'bagus', 'pembicara', 'dan', 'narasumbernya']

4.5.4. Normalization

Tahap ini bertujuan untuk menyamakan format penulisan kata supaya konsisten dan memudahkan analisis. Kata non-standar atau tidak baku dikonversi ke bentuk standar atau baku. Tabel 7. Merupakan hasil dari normalization.

Tabel 7. Hasil Normalization

Sebelum	Sesudah
['sbaiknya', 'jangan', 'di', 'hari', 'weekend']	['sebaiknya', 'jangan', 'di', 'hari', 'akhir pekan']

4.5.5. Stopword Removal

Tahap ini bertujuan untuk menghapus kata yang tidak penting atau terlalu umum seperti “dan”, “atau”, “yang”, “di”, dsb. Kata tersebut termasuk dalam *stopwords* yang diabaikan oleh mesin.

Tabel 8. Hasil Normalization

Sebelum	Sesudah
['sebaiknya', 'jangan', 'di', 'hari', 'akhir pekan']	['sebaiknya', 'jangan', 'hari', 'akhir pekan']

4.5.6. Topic Modeling

Tahap ini bertujuan untuk mengidentifikasi topik atau aspek dalam dokumen. Metode LDA diterapkan untuk mengidentifikasi topik utama dalam acara. Tahapan yang dilalui meliputi, mengubah format token ke bentuk list, membuat kamus menggunak corpora.Dictionary, lalu mengonversinya menjadi representasi *Bag of Words* (BoW) menggunakan *doc2bow()*, TF-DF (*Term Frequency-Inverse Document Frequency*) digunakan untuk memilih kata-kata yang lebih relevan dan unik, sehingga analisis lebih terfokus pada topik utama.

Model diuji dengan rentang topik 2 hingga 10, dan setiap model dievaluasi menggunakan metrik koherensi *'c_v'*. Berdasarkan Tabel 9, hasil coherence score menunjukkan bahwa 9 topik memperoleh skor koherensi tertinggi sebesar 0.707. Oleh karena itu, jumlah topik ini dipilih sebagai yang paling optimal untuk analisis topik.

Tabel 9. Hasil Coherence Score

Topic	Coherence	Topic	Coherence
2	0.373	7	0.650
3	0.503	8	0.681
4	0.581	9	0.707
5	0.641	10	0.706
6	0.687	-	-

Setelah menentukan jumlah topik optimal, analisis lebih lanjut terhadap topik dengan mengidentifikasi kata-kata berbobot tertinggi yang mencerminkan tema utama dari topik. Kata-kata ini digunakan untuk menamai setiap topik, sehingga memudahkan identifikasi aspek utama dalam acara seminar.

Tabel 10. Menyajikan hasil intrepetasi aspek yang diidentifkasi dari 7 topik. Penaaman aspek ini memberikan gambaran tentang tema atau aspek utama yang muncul dalam acara seminar.

Tabel 10. Intrepetasi Aspek

No	Intrepetasi Aspek
1	Manfaat atau Materi
2	Durasi atau Waktu
3	Kualitas Acara
4	Evaluasi atau Saran
5	Penyampaian Materi atau Kelancaran Acara
6	Antusiasme atau Apresiasi
7	Harapan Acara
8	Kesan atau Pesan
9	Narasumber atau Pembawa Acara

4.6. Klasifikasi Sentimen

Klasifikas sentimen dilakukan menggunakan *Support Vector Machine* (SVM). Sebelum pelatihan model, data teks dikonversi ke representasi numerik dengan TF-IDF untuk mencerminkan bobot kata berdasarkan frekuensi lokal dan global dalam dataset. Selanjutnya, berbagai skenario pengujian diterapkan untuk menentukan kombinasi terbaik dari metode *resampling*, dan pembagian data guna memperoleh akurasi optimal. Tabel 11. Merupakan hasil pengujian skenario SVM yang dilakukan.

Tabel 11. Hasil Pengujian SVM

Holdout		Akurasi		Precision	Recall	F1-Score
		Train	Test			
90:10	Normal	0.81	0.81	0.91	0.50	0.45
	SMOTE	0.89	0.85	0.78	0.88	0.81
80:20	Normal	0.81	0.82	0.91	0.50	0.45
	SMOTE	0.92	0.91	0.59	0.50	0.15

Berdasarkan hasil evaluasi model klasifikasi sentimen menggunakan SVM, penerapan SMOTE terbukti meningkatkan performa model secara signifikan, terutama dalam mengatasi ketidakseimbangan kelas pada dataset. Tanpa SMOTE, model mengalami *underfitting* dengan akurasi yang rendah sekitar 4-5%, terutama pada kelas minoritas, yang menunjukkan bahwa model gagal mengenali kelas tersebut dengan baik. Namun, setelah menerapkan SMOTE, terjadi peningkatan pada akurasi dan metrik evaluasi lainnya.

Skenario 90:10 dengan SMOTE menghasilkan performa terbaik dengan akurasi 85%, Precision 0.78, Recall 0.88, dan F1-Score 0.81, menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Sementara itu, skenario 80:20 dengan SMOTE mencapai akurasi lebih tinggi 91%, tetapi precision turun menjadi 0.59 dan F1-Score hanya 0.15, mengindikasikan potensi *overfitting* atau kurangnya generalisasi model terhadap data test.

Hasil ini menunjukkan bahwa meskipun SMOTE meningkatkan akurasi secara keseluruhan, pemilihan proporsi data latih dan uji juga mempengaruhi performa model. Untuk menghindari overfitting, diperlukan optimasi lebih lanjut seperti hyperparameter tuning atau pendekatan ensemble agar model dapat bekerja lebih optimal dalam skenario dunia nyata.

4.7. Analisis Sentimen Berbasis Aspek

Setelah topik acara seminar diidentifikasi dengan LDA, dan sentimen tiap topik diklasifikasikan menggunakan SVM. Analisis sentimen berbasis aspek menunjukkan distribusi sentimen negatif dan netral, yang dirangkum dalam Tabel 12. Distribusi jumlah sentimen positif dan negatif dalam berbagai aspek yang telah diidentifikasi dapat dilihat secara lebih rinci pada Tabel 12.

Tabel 12. Distribusi Sentimen Berbasis Aspek

Topic	Aspek	Positif	Negatif
1	Manfaat atau Materi	260	24
2	Durasi atau Waktu	39	42
3	Kualitas Acara	97	4
4	Evaluasi atau Saran	84	11
5	Penyampaian Materi atau Kelancaran Acara	43	17
6	Antusiasme atau Apresiasi	8	2
7	Harapan Acara	2	3
8	Kesan atau Pesan	1	6

Topic	Aspek	Positif	Negatif
9	Narasumber atau Pembawa Acara	13	1
Total Data		547	110

Berdasarkan Tabel 12, mayoritas peserta memberikan tanggapan positif terhadap acara, dengan 547 tanggapan positif (83.2%) dan 110 tanggapan negatif (16.8%). Aspek Manfaat atau Materi mendapatkan apresiasi tertinggi dengan 260 tanggapan positif dan hanya 24 tanggapan negatif, menunjukkan bahwa peserta merasa materi yang disampaikan sangat bermanfaat. Aspek Kualitas Acara dan Evaluasi atau Saran juga didominasi sentimen positif, masing-masing dengan 97 dan 84 tanggapan positif, yang mencerminkan kepuasan peserta terhadap penyelenggaraan seminar sekaligus adanya beberapa masukan untuk perbaikan. Selain itu, aspek Penyampaian Materi atau Kelancaran Acara juga memperoleh respons positif yang cukup besar (43 positif, 17 negatif), menunjukkan bahwa peserta umumnya puas dengan cara materi disampaikan, meskipun masih ada beberapa kritik terkait kelancaran acara.

Di sisi lain, aspek Durasi atau Waktu memiliki sentimen yang lebih seimbang (39 positif, 42 negatif), menunjukkan adanya ketidakpuasan terkait jadwal atau durasi acara. Aspek Kesan atau Pesan juga memiliki kecenderungan negatif dengan 6 tanggapan negatif dan hanya 1 tanggapan positif, yang dapat mengindikasikan adanya pengalaman kurang menyenangkan dari beberapa peserta. Sementara itu, aspek Antusiasme atau Apresiasi (8 positif, 2 negatif) dan Harapan Acara (2 positif, 3 negatif) mendapat jumlah tanggapan yang relatif sedikit, menunjukkan bahwa topik ini bukan menjadi perhatian utama peserta.

Aspek Narasumber atau Pembawa Acara memperoleh penilaian yang mayoritas positif (13 positif, 1 negatif), mengindikasikan bahwa peserta umumnya puas dengan performa pembicara dalam seminar ini. Secara keseluruhan, acara seminar mendapatkan tanggapan yang sangat positif, terutama dari segi manfaat materi, kualitas acara, dan penyampaian materi. Namun, terdapat beberapa kritik terhadap durasi acara serta kesan yang ditinggalkan, yang dapat menjadi perhatian untuk perbaikan dalam pelaksanaan seminar berikutnya.

5. KESIMPULAN DAN SARAN

Berdasarkan analisis sentimen berbasis aspek terhadap respons survei *open-ended* acara Seminar Nasional FKBM-IK 2021 dan 2022, metode LDA berhasil mengidentifikasi 9 topik utama, sementara klasifikasi sentimen menggunakan SVM dengan skenario 90:10 dan SMOTE mencapai performa terbaik dengan akurasi 85%, precision 0.78, recall 0.88, dan F1-score 0.81. Hasil ini menunjukkan pendekatan yang cukup optimal, namun masih dapat ditingkatkan. Saran untuk penelitian selanjutnya dapat mempertimbangkan metode yang lebih canggih untuk meningkatkan performa model serta analisis lebih dalam terhadap faktor yang memengaruhi sentimen negatif, guna memberikan rekomendasi yang lebih akurat bagi perbaikan acara di masa mendatang.

DAFTAR PUSTAKA

[1] H. Muhammad, “Inovasi Platform Survey Online Bagi Layanan Informasi Bisnis Masa Kini,” Republika Online, Apr. 27, 2022. [Online]. Available:https://ekonomi.republika.co.id/berita/raydes380/inovasi-platform-surveyonline-bagi-layanan-informasi-bisnis-masa-kini

[2] BuiltWith, “Feedback Forms and Surveys Usage Distribution in Indonesia,” BuiltWith. Accessed: Dec. 18, 2024. [Online]. Available: https://trends.builtwith.com/widgets/feedback-forms-andsurveys/country/Indonesia

[3] N. Nikolić, O. Grljević, and A. Kovačević, “Aspect-based sentiment analysis of reviews in the domain of higher education,” The Electronic Library, vol. 38, no. 1, pp. 44–64, Feb. 2020, doi: 10.1108/el-06-2019-0140.

[4] V. Baburajan, J. de A. e Silva, and F. C. Pereira, “Open-Ended Versus Closed-Ended Responses: A Comparison Study Using Topic Modeling and Factor Analysis,” IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 4, pp. 2123–2132, Apr. 2021, doi: 10.1109/tits.2020.3040904.

[5] J. Rouder, O. Saucier, R. Kinder, and M. Jans, “What to Do With All Those Open-Ended Responses? Data Visualization Techniques for Survey Researchers,” Survey Practice, vol. 14, no. 1, pp. 1–9, Aug. 2021, doi: 10.29115/sp-2021-0008.

[6] H. Liu, I. Chatterjee, M. Zhou, X. S. Lu, and A. Abusorrah, “Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods,” IEEE Transactions on Computational Social Systems, vol. 7, no. 6, pp. 1358–1375, Dec. 2020, doi: 10.1109/tcss.2020.3033302.

[7] S. Poria, D. Hazarika, N. Majumder, and R. Mihalcea, “Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research,” IEEE Transactions on Affective Computing, vol. 14, no. 1, pp. 108–132, Jan. 2023, doi: 10.1109/taffc.2020.3038167.

[8] M. D. R Wahyudi, A. Fatwanto, U. Kiftiyani, and M. Galih Wonoseto, “Topic Modeling of Online Media News Titles during COVID-19 Emergency Response in Indonesia Using the Latent Dirichlet Allocation (LDA) Algorithm,” Telematika, vol. 14, no. 2, pp. 101–111, Aug. 2021, doi: 10.35671/telematika.v14i2.1225.

[9] P. Andono Nurtantio, Sunardi, R. Arief Nugroho, and B. Harjo, “Aspect-Based Sentiment Analysis for Hotel Review Using LDA, Semantic Similarity, and BERT,” *International Journal of*

Intelligent Engineering and Systems, vol. 15, no. 5, pp. 232–243, Oct. 2022, doi: 10.22266/ijies2022.1031.21

[10] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," Neurocomputing, vol. 408, pp. 189–215, Sep. 2020, doi: 10.1016/j.neucom.2019.10.118.

[11] Y. Kustiyahningsih and Y. Permana, "Penggunaan Latent Dirichlet Allocation (LDA) dan Support-Vector Machine (SVM) Untuk Menganalisis Sentimen Berdasarkan Aspek Dalam Ulasan Aplikasi EdLink," Teknika, vol. 13, no. 1, pp. 127–136, Mar. 2024, doi: 10.34148/teknika.v13i1.746.

[12] M. A. Palimbani, R. P. Hastuti, and R. A. Rajagede, "Analisis Sentimen Berbasis Aspek Pada Ulasan Pengguna Aplikasi Starbucks Menggunakan Algoritma Support Vector Machine," Journal of Internet and Software Engineering, vol. 5, no. 1, pp. 43–49, May 2024, doi: 10.22146/jise.v5i1.9130..

[13] A. A. Arifiyanti, D. S. Y. Kartika, and C. J. Prawiro, "Using Pre-Trained Models for Sentiment Analysis in Indonesian Tweets," unknown.[Online].Available:https://www.researchgate.net/publication/365103819_Using_Pre-Trained_Models_for_Sentiment_Analysis_in_Indonesian_Tweets.

[14] A. I. Alfanzar, K. Khalid, and I. S. Rozas, "TOPIC MODELLING SKRIPSI MENGGUNAKAN METODE LATENT DIRICLHET ALLOCATION," *JSiI (Jurnal Sistem Informasi)*, vol. 7, no. 1, p. 7, Mar. 2020, doi: 10.30656/jsii.v7i1.2036.

[15] R. Kusnadi, Y. Yusuf, A. Andriantony, and M. Caintan, "ANALISIS SENTIMEN TERHADAP GAME GENSHIN IMPACT MENGGUNAKAN BERT," LPPM Universitas Abdurrab. [Online]. Available:https://www.researchgate.net/publication/353176928_ANALISIS_SENTIMEN_TERHADAP_GAME_GENSHIN_IMPACT_MENGGUNAKAN_BERT