



Analisis Sentimen Dialog Film Spider-Man Series Menggunakan Complement Naïve Bayes

Salma Faradhiba^{1*}, Preity Deldiana Putri², Aryo Nugroho³

^{1*3} Program Studi Sistem Informasi, Universitas Narotama, Indonesia

Alamat : Jl. Arief Rahman Hakim No.51, Sukolilo, Kota Surabaya, Jawa Timur.

Email: ¹salmafaradhiba08@gmail.com, ²preitydel@gmail.com, ³aryo.nugroho@narotama.ac.id

Penulis Korespondensi: salmafaradhiba08@gmail.com

Abstract: The rapid growth of the film industry and the wider adoption of Natural Language Processing (NLP) have increased the importance of analyzing film dialogues to identify thematic patterns. This study classifies dialogues from Tom Holland's Spider-Man trilogy into three thematic categories: Conflict, Friendship, and Multiverse, using the Complement Naive Bayes algorithm. The dataset was generated through audio extraction, Speech-to-Text transcription, and verification with the original dialogue scripts. The methodology included text preprocessing, keyword-based automatic labeling, TF-IDF feature extraction, Complement Naive Bayes classification with an 80:20 training-testing split, accuracy evaluation, and Word Cloud visualization. The proposed model achieved accuracies of 84% for Spider-Man: Homecoming, 77% for Spider-Man: Far From Home, and 94% for Spider-Man: No Way Home. The Word Cloud results identified dominant words representing the thematic characteristics of each film. Overall, the combination of TF-IDF and Complement Naive Bayes effectively classifies film dialogues and demonstrates its potential for film script analysis and content-based recommendation systems.

Keywords: Sentiment analysis; Film dialogue; Complement Naive Bayes; TF-IDF; Theme classification.

Abstrak: Perkembangan industri film yang diiringi dengan semakin luasnya penerapan Natural Language Processing (NLP) mendorong pemanfaatan dialog film sebagai sumber data tekstual untuk mengidentifikasi tema dan pola bahasa. Penelitian ini bertujuan mengklasifikasikan dialog pada trilogi film Spider-Man versi Tom Holland ke dalam tiga kategori tematik, yaitu Konflik, Persahabatan, dan Multiverse, menggunakan algoritma Complement Naive Bayes. Dataset disusun melalui proses ekstraksi audio, konversi Speech-to-Text, serta verifikasi menggunakan naskah dialog asli. Tahapan penelitian meliputi *preprocessing* (*text cleaning*, *tokenization*, *stopword removal*, dan *stemming*), pelabelan otomatis berbasis keyword, ekstraksi fitur menggunakan TF-IDF, pemodelan Complement Naive Bayes dengan pembagian data latih dan data uji sebesar 80:20, evaluasi menggunakan *accuracy*, serta visualisasi Word Cloud. Hasil penelitian menunjukkan akurasi sebesar 84% pada *Spider-Man: Homecoming*, 77% pada *Spider-Man: Far From Home*, dan 94% pada *Spider-Man: No Way Home*. Visualisasi Word Cloud juga menampilkan kata-kata dominan yang merepresentasikan karakteristik tema pada setiap film. Temuan ini menunjukkan bahwa kombinasi TF-IDF dan Complement Naive Bayes efektif dalam mengklasifikasikan dialog film berdasarkan kategori tema serta berpotensi diterapkan pada analisis naskah film dan pengembangan sistem rekomendasi berbasis konten.

Kata kunci: Analisis sentimen; Dialog film; Complement Naive Bayes; TF-IDF; Klasifikasi tema.

1. LATAR BELAKANG

Perkembangan teknologi digital telah mendorong pertumbuhan industri film sebagai media hiburan sekaligus sarana penyampaian informasi, emosi, dan nilai sosial. Dialog film berperan penting dalam membangun alur cerita dan mengembangkan karakter sehingga menjadi sumber data tekstual yang potensial untuk dianalisis menggunakan Natural Language Processing (NLP) (Mordor Intelligence, 2024;

Naskah Masuk: 15 Agustus 2026; Revisi: 20 Agustus 2026; Diterima: 28 September 2026; Tersedia: 10 September 2026; Terbit: 12 Oktober 2026

Pracintya et al., 2024; Rachim, 2023). Salah satu penerapan NLP adalah analisis sentimen yang digunakan untuk mengidentifikasi kecenderungan informasi pada data teks melalui proses klasifikasi guna mendukung analisis dan pengambilan keputusan (Mursyid & Indriyanti, 2024). Trilogi Spider-Man versi Tom Holland dipilih sebagai objek penelitian karena dialognya memuat berbagai tema, seperti konflik, persahabatan, dan multiverse, yang kaya akan konteks naratif serta ekspresi emosional (Ergün et al., 2024; Hanmei, 2024).

Penelitian terdahulu umumnya menerapkan analisis sentimen pada ulasan film maupun media sosial menggunakan algoritma seperti Naïve Bayes dan Support Vector Machine dengan hasil yang baik (Cahya Kamilla et al., 2024; Fathullah et al., 2020; Gifari et al., 2022). Namun, penelitian tersebut masih berfokus pada opini pengguna, sedangkan pemanfaatan dialog film sebagai sumber data utama, khususnya menggunakan algoritma Complement Naïve Bayes untuk klasifikasi kategori tematik, masih terbatas (Ergün et al., 2024; Hanmei, 2024).

Berdasarkan kesenjangan penelitian tersebut, algoritma Complement Naïve Bayes diterapkan dalam penelitian ini untuk mengklasifikasikan dialog film ke dalam tiga kategori tematik, yaitu Konflik, Persahabatan, dan Multiverse. Algoritma ini dipilih karena dirancang untuk meningkatkan performa klasifikasi data teks, terutama pada distribusi kelas yang tidak seimbang (Budiman et al., 2021). Penelitian ini diharapkan dapat memperluas penerapan NLP pada analisis dialog film serta memberikan alternatif penerapan Complement Naïve Bayes dalam klasifikasi teks bertema.

2. KAJIAN TEORITIS

Natural Language Processing (NLP) merupakan salah satu bidang dalam *Artificial Intelligence* (AI) yang berfokus pada pengembangan metode agar komputer mampu memahami, mengolah, dan menganalisis bahasa alami manusia. Salah satu implementasi NLP yang banyak digunakan adalah analisis sentimen, yaitu teknik yang bertujuan mengidentifikasi serta mengklasifikasikan opini, emosi, atau kecenderungan suatu teks berdasarkan informasi yang terkandung di dalamnya (Mursyid & Indriyanti, 2024). Penerapan analisis sentimen telah berkembang pada berbagai jenis data tekstual, termasuk dialog film yang memiliki alur bahasa lebih terstruktur dan konteks naratif yang konsisten. Karakteristik tersebut memungkinkan dialog film merepresentasikan interaksi

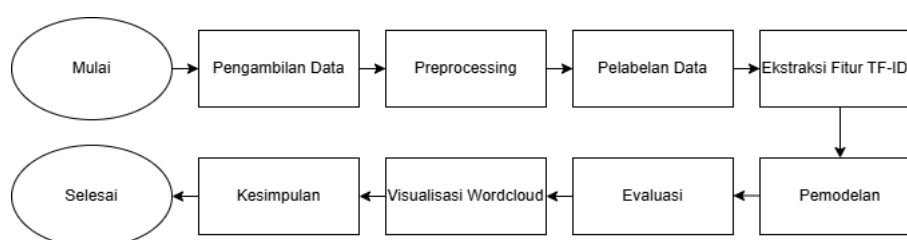
antartokoh beserta ekspresi emosional secara lebih kontekstual (Ergün et al., 2024; Hanmei, 2024). Pada penelitian ini, proses klasifikasi dilakukan menggunakan algoritma Complement Naïve Bayes (CNB), yang merupakan pengembangan dari Multinomial Naïve Bayes. Algoritma ini dirancang untuk meningkatkan kinerja klasifikasi, terutama pada data teks dengan distribusi kelas yang tidak seimbang (*imbalanced dataset*), melalui pemanfaatan bobot yang dihitung berdasarkan kelas komplemen sehingga dapat mengurangi kecenderungan bias terhadap kelas mayoritas (Budiman et al., 2021). Persamaan matematis yang mendasari algoritma Complement Naïve Bayes disajikan pada Persamaan (P1).

$$score(c, d) = \sum_{i=1}^n x_i \cdot \log \left(\frac{P(w_i|\bar{c})}{\sum_j P(w_j|\bar{c})} \right) - \log(P(c)) \quad (P1)$$

3. METODE PENELITIAN

3.1 Tahapan Penelitian

Penelitian ini menggunakan bahasa pemrograman Python sebagai platform utama dalam melaksanakan seluruh tahapan analisis. Proses penelitian diawali dengan pengambilan data, kemudian dilanjutkan dengan *preprocessing*, pelabelan data berbasis *keyword*, ekstraksi fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), serta pemodelan menggunakan algoritma Complement Naïve Bayes. Selanjutnya, kinerja model dievaluasi menggunakan metrik *accuracy*, kemudian hasil klasifikasi divisualisasikan dalam bentuk *Word Cloud* sebelum dilakukan penarikan kesimpulan. Tahapan penelitian secara keseluruhan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

3.2 Pengambilan Data

Data penelitian berupa dialog dari tiga film Spider-Man versi Tom Holland, yaitu *Spider-Man: Homecoming* (2017), *Spider-Man: Far From Home* (2019), dan *Spider-Man: No Way Home* (2021). Audio dari masing-masing film diekstraksi menggunakan

bahasa pemrograman Python, kemudian dikonversi menjadi teks melalui proses *Speech-to-Text* (Shafirra & Irhamah, 2020). Hasil konversi selanjutnya diverifikasi dan disesuaikan dengan naskah dialog (script) untuk memperbaiki ketidaksesuaian akibat kesalahan pengenalan suara sebelum disusun menjadi dataset penelitian.

3.3 Preprocessing

Tahap *preprocessing* dilakukan menggunakan bahasa pemrograman Python untuk menghasilkan data teks yang bersih dan konsisten sebelum proses pelabelan serta klasifikasi. Tahapan ini meliputi *text cleaning*, *tokenization*, *stopword removal*, dan *stemming* dengan bantuan pustaka NLTK. Proses tersebut bertujuan menghilangkan informasi yang tidak relevan, memisahkan teks menjadi token, menghapus *stopword*, serta mengubah kata ke bentuk dasar agar kualitas data lebih optimal untuk tahap analisis berikutnya (Karimah, 2024; Putra Negara, 2023; Santosa et al., 2022; Umar & M. Adnan Nur, 2022).

3.4 Pelabelan Data

Pelabelan data dilakukan secara otomatis menggunakan bahasa pemrograman Python berdasarkan kumpulan *keyword* yang disusun dari dialog trilogi film Spider-Man. *Keyword* tersebut dikelompokkan ke dalam tiga kategori tematik, yaitu Konflik, Persahabatan, dan Multiverse. Setiap dialog kemudian dicocokkan dengan *keyword* pada masing-masing kategori, dan label ditentukan berdasarkan jumlah kecocokan tertinggi. Dialog yang tidak memiliki kecocokan pada seluruh kategori diberi label "Tidak Diketahui". Label yang dihasilkan selanjutnya digunakan sebagai *ground truth* pada proses ekstraksi fitur TF-IDF dan klasifikasi menggunakan algoritma Complement Naïve Bayes.

3.5 Ekstraksi Fitur TF-IDF

Setelah pelabelan data, teks dikonversi ke dalam representasi numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan tingkat kemunculannya pada seluruh dokumen, sehingga dapat merepresentasikan tingkat kepentingan setiap kata. Selanjutnya, hasil pembobotan TF-IDF digunakan sebagai fitur masukan dalam proses klasifikasi menggunakan algoritma Complement Naïve Bayes (Aryanti & Mahendra, 2023; Br Sinulingga & Sitorus, 2024).

3.6 Pemodelan

Proses klasifikasi dilakukan menggunakan algoritma Complement Naïve Bayes (CNB) dengan memanfaatkan fitur hasil pembobotan TF-IDF. Dataset dibagi menjadi 80% data latih dan 20% data uji. Data latih digunakan untuk membangun model, sedangkan data uji berfungsi mengevaluasi kemampuan model dalam mengklasifikasikan dialog ke dalam kategori Konflik, Persahabatan, dan Multiverse (Budiman et al., 2021; Putra & Nugroho, 2021).

3.7 Evaluasi

Evaluasi model dilakukan menggunakan metrik accuracy untuk mengukur ketepatan algoritma Complement Naïve Bayes dalam mengklasifikasikan dialog ke dalam kategori Konflik, Persahabatan, dan Multiverse. Nilai accuracy dihitung berdasarkan perbandingan jumlah prediksi yang benar terhadap keseluruhan data uji. Perhitungan accuracy mengacu pada Persamaan (P2) (Prayogo et al., 2023).

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Total\ Data\ Uji} \times 100\% \quad (P2)$$

3.8 Visualisasi Word Cloud

Word Cloud digunakan untuk menampilkan kata-kata yang paling dominan berdasarkan frekuensi kemunculannya dalam dataset (Pradana, 2020). Pada penelitian ini, *Word Cloud* dibuat menggunakan bahasa pemrograman Python untuk memvisualisasikan kata-kata dominan pada setiap kategori, yaitu Konflik, Persahabatan, dan Multiverse. Visualisasi tersebut digunakan sebagai pendukung interpretasi terhadap hasil klasifikasi sehingga karakteristik kata pada masing-masing kategori dapat diamati dengan lebih mudah.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pengambilan Data

Proses pengambilan data menghasilkan dataset dialog dari tiga film Spider-Man versi Tom Holland, yaitu *Spider-Man: Homecoming* (2017), *Spider-Man: Far From Home* (2019) dan *Spider-Man: No Way Home* (2021). Dataset diperoleh melalui ekstraksi audio, proses *Speech-to-Text*, serta validasi menggunakan script dialog untuk memperbaiki ketidaksesuaian hasil transkripsi. Hasil pengambilan data selanjutnya digunakan sebagai dataset pada tahap *preprocessing*.

4.2 Hasil Preprocessing

Tahap *preprocessing* menghasilkan data teks yang telah dibersihkan dan diseragamkan sehingga siap digunakan pada proses pelabelan serta ekstraksi fitur TF-IDF. Setelah melalui proses *text cleaning*, *tokenization*, *stopword removal*, dan *stemming*, setiap dialog telah direpresentasikan dalam bentuk kata dasar dengan karakter yang tidak diperlukan dihilangkan. Hasil *preprocessing* selanjutnya digunakan sebagai masukan pada proses pelabelan data. Contoh hasil *preprocessing* ditunjukkan pada Tabel 1.

Tabel 1. *Preprocessing*

Tahap	Hasil
Data Awal	Provided by explosiveskull We come
<i>Text Cleaning</i>	provided by explosiveskull we come
<i>Tokenization</i>	['provided', 'by', 'explosiveskull', 'we', 'come']
<i>Stopword</i>	['provided', 'explosiveskull', 'come']
<i>Stemming</i>	['provid', 'explosiveskul', 'come']

4.3 Hasil Pelabelan Data

Proses pelabelan otomatis menggunakan bahasa pemrograman Python berhasil mengelompokkan dialog ke dalam kategori Konflik, Persahabatan, dan Multiverse berdasarkan kecocokan dengan daftar keyword yang telah ditentukan. Hasil pelabelan menghasilkan dataset berlabel yang digunakan sebagai ground truth dalam proses pelatihan dan pengujian model Complement Naïve Bayes. Dataset tersebut selanjutnya menjadi dasar pada tahap ekstraksi fitur TF-IDF dan proses klasifikasi.

4.4 Hasil Ekstraksi Fitur TF-IDF

Tahap ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) menghasilkan representasi numerik sebagai masukan bagi algoritma Complement Naïve Bayes. Selain itu, proses ini juga mengidentifikasi kata-kata dominan pada dialog setiap film. Tabel 2 menyajikan sepuluh kata dominan hasil ekstraksi fitur TF-IDF beserta frekuensi kemunculannya pada masing-masing film.

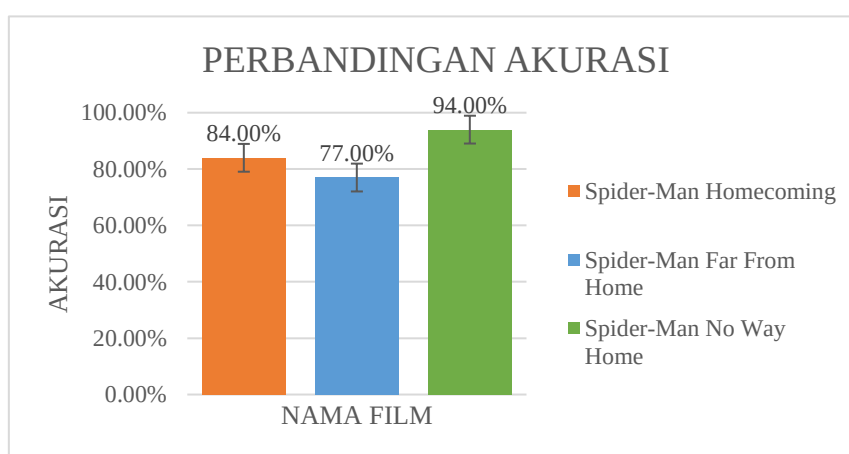
Tabel 2. Representasi TF-IDF

Nama Film	10 Kata Tertinggi TF-IDF	Frekuensi Kata
<i>Spider-Man: Homecoming</i>	<i>great, friend, kill, help, understand, fun, love, hope, team, fight</i>	13, 12, 11, 9, 9, 8, 7, 7, 6, 5.
<i>Spider-Man: Far From Home</i>	<i>drone, kill, help, great, friend, love, save, hit, die, danger</i>	19, 16, 15, 14, 12, 10, 10, 8, 8, 7.
<i>Spider-Man: No Way Home</i>	<i>help, spell, die, kill, love, friend, great, power, cure, care</i>	39, 22, 18, 18, 17, 16, 13, 11, 10, 8.

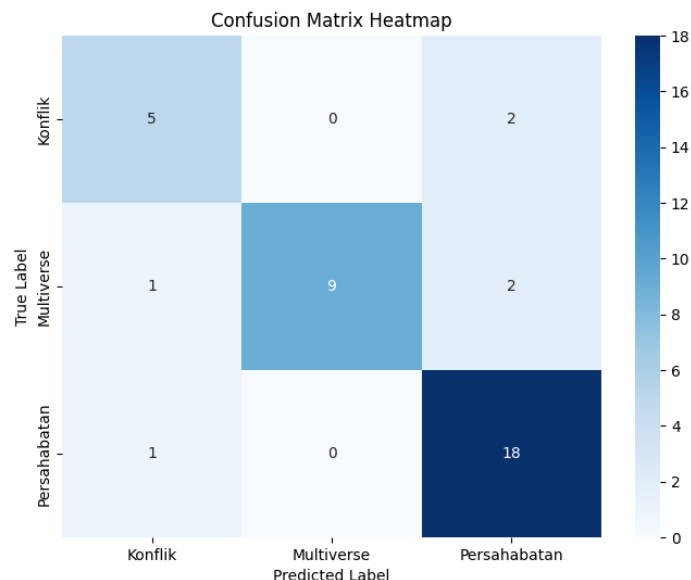
Berdasarkan Tabel 2, setiap film memiliki kata-kata dominan yang berbeda sesuai dengan konteks cerita yang disampaikan. Pada *Spider-Man: Homecoming*, kata *great* dan *friend* memiliki frekuensi kemunculan tertinggi. Sementara itu, *Spider-Man: Far From Home* didominasi oleh kata *drone*, yang berkaitan dengan konflik utama dalam cerita, sedangkan *Spider-Man: No Way Home* didominasi oleh kata *help* dan *spell*, yang mencerminkan tema multiverse dan peran sihir dalam alur cerita. Perbedaan kata-kata dominan tersebut menunjukkan bahwa ekstraksi fitur TF-IDF mampu merepresentasikan karakteristik dialog pada masing-masing film sehingga dapat mendukung proses klasifikasi menggunakan algoritma Complement Naïve Bayes.

4.5 Hasil Klasifikasi Complement Naïve Bayes

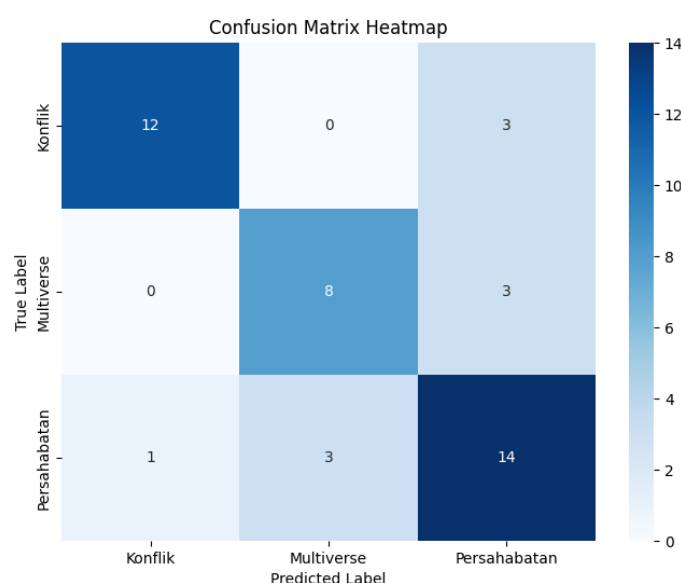
Model Complement Naïve Bayes digunakan untuk mengklasifikasikan dialog ke dalam kategori Konflik, Persahabatan, dan Multiverse berdasarkan representasi fitur TF-IDF. Kinerja model pada setiap dataset kemudian dievaluasi menggunakan metrik accuracy untuk mengukur tingkat ketepatan klasifikasi.

**Gambar 2.** Perbandingan Akurasi Klasifikasi pada Masing-masing Film Spider-Man

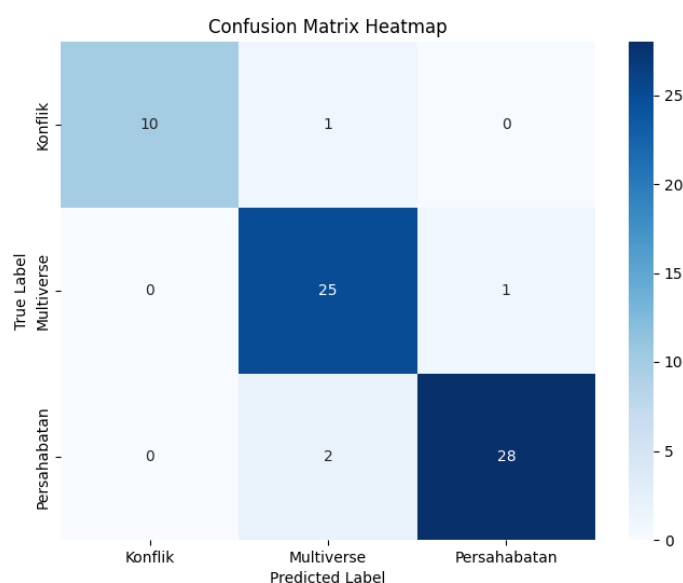
Berdasarkan Gambar 2, hasil pengujian menunjukkan bahwa model Complement Naïve Bayes memperoleh tingkat akurasi sebesar 84% pada *Spider-Man: Homecoming*, 77% pada *Spider-Man: Far From Home*, dan 94% pada *Spider-Man: No Way Home*. Tingginya akurasi pada No Way Home menunjukkan bahwa karakteristik kosakata pada setiap kategori tematik relatif lebih konsisten sehingga model lebih mudah membedakan masing-masing kelas. Sebaliknya, akurasi Far From Home merupakan yang terendah karena dialog dalam film tersebut memiliki konteks yang lebih beragam, di mana unsur konflik, persahabatan, dan tanggung jawab sering muncul secara bersamaan sehingga meningkatkan kemiripan fitur antarkategori. Sementara itu, No Way Home memperoleh akurasi yang lebih tinggi dibandingkan Far From Home karena tema multiverse menjadi fokus utama cerita, sehingga banyak dialog mengandung kosakata yang lebih spesifik, seperti *spell*, *multiverse*, dan karakter dari semesta lain, yang membantu model mengenali perbedaan antar kategori. Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma Complement Naïve Bayes mampu mengklasifikasikan dialog film secara efektif sesuai dengan kategori tematik yang telah ditentukan. Selanjutnya, hasil klasifikasi divisualisasikan menggunakan *confusion matrix* untuk memperlihatkan kesesuaian antara label aktual dan hasil prediksi model. Hasil *confusion matrix* pada masing-masing film ditunjukkan pada Gambar 3, Gambar 4 dan Gambar 5.



Gambar 3. Confusion Matrix Hasil Klasifikasi pada *Spider-Man Homecoming*



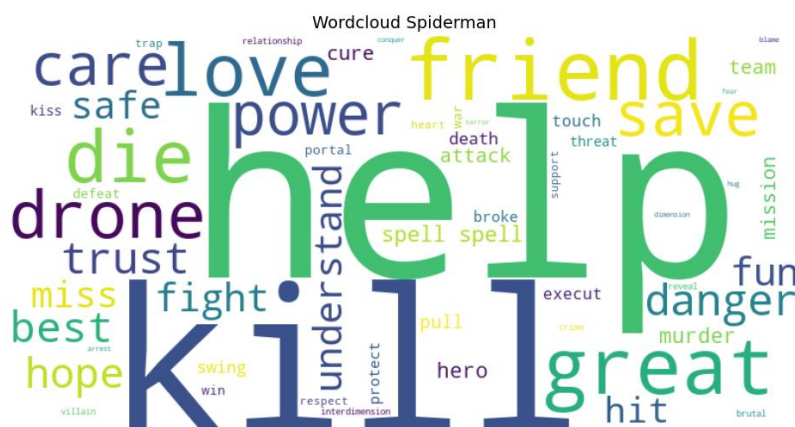
Gambar 4. Confusion Matrix Hasil Klasifikasi pada *Spider-Man Far From Home*



Gambar 5. Confusion Matrix Hasil Klasifikasi pada *Spider-Man No Way Home*

Berdasarkan Gambar 3, Gambar 4 dan Gambar 5 menampilkan *confusion matrix* hasil klasifikasi pada ketiga film Spider-Man. Secara umum, nilai pada diagonal utama matriks lebih dominan dibandingkan nilai di luar diagonal, yang menunjukkan bahwa sebagian besar data berhasil diklasifikasikan sesuai dengan label sebenarnya. Meskipun masih terdapat beberapa kesalahan klasifikasi, hasil tersebut menunjukkan bahwa algoritma Complement Naïve Bayes mampu mengklasifikasikan tema dialog dengan baik pada ketiga film.

4.6 Visualisasi *Word Cloud*



Berdasarkan Gambar 6, visualisasi *Word Cloud* menampilkan kata-kata yang paling sering muncul dalam dialog trilogi Spider-Man. Kata seperti Peter, *know*, *yeah*, *go*, *look*, *help*, *spell*, dan *die* memiliki frekuensi kemunculan yang tinggi karena berkaitan dengan perkembangan alur cerita dan interaksi antartokoh. Kata *help* merepresentasikan upaya Peter Parker dalam membantu orang lain maupun para karakter dari semesta lain, sedangkan kata *spell* menjadi salah satu kata dominan karena berkaitan dengan mantra yang digunakan Doctor Strange dalam *Spider-Man: No Way Home*. Sementara itu, kata *die* muncul sebagai bagian dari konflik mengenai takdir beberapa karakter dari semesta lain. Dominasi kata-kata tersebut menunjukkan bahwa dialog dalam trilogi Spider-Man tidak hanya berisi percakapan sehari-hari, tetapi juga merefleksikan konflik, hubungan antartokoh, serta konsep *multiverse* yang menjadi fokus utama cerita sehingga mendukung pembentukan kategori tematik pada penelitian ini.

Hasil penelitian menunjukkan bahwa algoritma Complement Naïve Bayes memiliki kemampuan yang baik dalam mengklasifikasikan dialog film ke dalam kategori Konflik, Persahabatan, dan Multiverse. Kinerja tersebut didukung oleh ekstraksi fitur TF-IDF yang

mampu merepresentasikan karakteristik kata dalam setiap dialog, sehingga memudahkan proses klasifikasi. Selain itu, mekanisme Complement Naïve Bayes yang memanfaatkan distribusi kelas komplemen membantu meningkatkan performa klasifikasi pada data teks (Budiman et al., 2021).

Perbedaan tingkat akurasi pada masing-masing film menunjukkan bahwa karakteristik dialog, seperti variasi kosakata dan konteks percakapan, memengaruhi kemampuan model dalam melakukan klasifikasi. Hasil penelitian ini sejalan dengan penelitian (Ergün et al., 2024) dan (Hanmei, 2024), yang menunjukkan bahwa teknik *text mining* dan analisis teks efektif dalam mengidentifikasi pola bahasa pada dialog. Dengan demikian, kombinasi TF-IDF dan Complement Naïve Bayes terbukti efektif untuk mengklasifikasikan tema dialog film.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menerapkan *algoritma Complement Naïve Bayes* untuk mengklasifikasikan dialog film Spider-Man ke dalam tiga kategori tematik, yaitu Konflik, Persahabatan, dan Multiverse. Hasil penelitian menunjukkan bahwa kombinasi ekstraksi fitur TF-IDF dan algoritma Complement Naïve Bayes mampu mengklasifikasikan tema dialog secara efektif berdasarkan karakteristik teks. Selain itu, visualisasi *Word Cloud* berhasil mengidentifikasi kata-kata dominan yang mendukung interpretasi karakteristik dialog pada dataset.

Penelitian selanjutnya disarankan menggunakan dataset yang lebih beragam serta membandingkan kinerja *Complement Naïve Bayes* dengan algoritma klasifikasi lain untuk memperoleh hasil evaluasi yang lebih komprehensif.

DAFTAR REFERENSI

- Aryanti, P. A. N., & Mahendra, I. B. M. (2023). Analisis Sentimen Opini Berbahasa Indonesia Pada Sosial Media Menggunakan TF-IDF dan Support Vector Machine. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 12(1), 45–52. <https://doi.org/10.24843/JLK.2023.v12.i01>
- Br Sinulingga, J. E., & Sitorus, H. C. K. (2024). Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF. *Jurnal Manajemen Informatika (JAMIKA)*, 14(1), 42–53. <https://doi.org/10.34010/jamika.v14i1.11946>

- Budiman, A., Young, J. C., & Suryadibrata, A. (2021). Implementasi Algoritma Naïve Bayes untuk Klasifikasi Konten Twitter dengan Indikasi Depresi. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(2), 133–138. <https://doi.org/10.30591/jpit.v6i2.2419>
- Cahya Kamilla, A., Priyani, N., Priskila, R., & Handrianus Pranatawijaya, V. (2024). Analisis Sentimen Film Agak Laen Dengan Kecerdasan Buatan: Text Mining Metode Naïve Bayes Classifier. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 2923–2928. <https://doi.org/10.36040/jati.v8i3.9587>
- Ergün, D., Takan, S., & Katipoğlu, G. (2024). Emotional Structure of Dialogues in Aging-Themed Movies: An Analysis on Text Mining. *Sage Open*, 14(3), 21582440241260502. <https://doi.org/10.1177/21582440241260502>
- Fathullah, N. S., Sari, Y. A., & Adikara, P. P. (2020). Analisis Sentimen Terhadap Rating dan Ulasan Film dengan menggunakan Metode Klasifikasi Naïve Bayes dengan Fitur Lexicon-Based. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(2), 590–593.
- Gifari, O. I., Adha, Muh., Freddy, F., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40. <https://doi.org/10.46229/jifotech.v2i1.330>
- Hanmei, W. (2024). Deep Learning-Based Analysis of Emotional and Content Relevance Between Bullet Screens and Subtitles as Movie Narrative Medium. *Sage Open*, 14(3), 1–19. <https://doi.org/10.1177/21582440241280840>
- Karimah, N. (2024). Multi-Aspect Sentiment Analysis Pada Review Film Menggunakan Metode Bidirectional Encoder Representations From Transformers (BERT). *Komputika : Jurnal Sistem Komputer*, 13(1), 63–72. <https://doi.org/10.34010/komputika.v13i1.11098>
- Mordor Intelligence. (2024). *Indonesia Digital Media Market Size & Share Analysis—Growth Trends & Forecasts (2025—2030)* [Situs Riset Pasar dan Konsultan Industri]. <https://www.mordorintelligence.com/industry-reports/indonesia-digital-media-market>
- Mursyid, R., & Indriyanti, A. D. (2024). Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter). *Journal of Informatics and Computer Science (JINACS)*, 06(02), 371–383. <https://doi.org/10.26740/jinacs.v6n02>
- Pracintya, I. A. K. D. W., Putrayasa, I. B., & Martha, I. N. (2024). Ragam Ekspresi Fisik Yang Ditunjukkan Dalam Dialog Film Ketika Berhenti Di Sini Karya Umay

- Shahab. *Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 14(4), 584–594.
<https://doi.org/10.23887/jpbsi.v14i4.93907>
- Pradana, M. G. (2020). Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam Text Mining. *Jurnal Ilmiah Informatika (JIF)*, 08(01), 38–43.
<https://doi.org/10.33884/jif.v8i01.1838>
- Prayogo, A., Fauziah, F., & Winarsih, W. (2023). Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor Pada Klasifikasi Judul Artikel Pada Jurnal Ilmiah. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 8(4), 1327–1338.
<https://doi.org/10.29100/jupi.v8i4.4141>
- Putra Negara, A. B. (2023). The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 14(3), 172.
<https://doi.org/10.24843/LKJITI.2023.v14.i03.p05>
- Putra, Z. P., & Nugroho, A. (2021). Pebandingan Performa Naïve Bayes dan KNN pada Klasifikasi Teks Sentimen Jasa Ekspedisi. *JOINTECS (Journal of Information Technology and Computer Science)*, 6(3), 145.
<https://doi.org/10.31328/jointecs.v6i3.2635>
- Rachim, H. I. (2023). Analisis Tindak Tutur Ilokusi Dalam Film Pendek Evakuasi Mama Emola. *Sabda: Jurnal Bahasa dan Sastra*, 1(2), 20–30.
<https://doi.org/10.572349/sabda.v1i2.358>
- Santosa, A., Purnamasari, I., & Mayasari, R. (2022). Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma LSTM. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 6(1), 81–93. <http://dx.doi.org/10.30645/j-sakti.v6i1.427>
- Shafirra, N. A., & Irhamah, I. (2020). Klasifikasi Sentimen Ulasan Film Indonesia dengan Konversi Speech-to-Text (STT) Menggunakan Metode Convolutional Neural Network (CNN). *Jurnal Sains dan Seni ITS*, 9(1), D95–D101.
<https://doi.org/10.12962/j23373520.v9i1.51825>
- Umar, N. & M. Adnan Nur. (2022). Application of Naïve Bayes Algorithm Variations On Indonesian General Analysis Dataset for Sentiment Analysis. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(4), 585–590.
<https://doi.org/10.29207/resti.v6i4.4179>