

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Pengambilan Data

Proses pengambilan data dalam penelitian ini menghasilkan kumpulan teks dialog yang bersumber dari tiga film dalam seri *Spider-Man*, yakni *Spider-Man: Homecoming* (2017), *Spider-Man: Far From Home* (2019), dan *Spider-Man: No Way Home* (2021). Ketiga film tersebut dianalisis dalam format video digital (.mp4) dengan audio berbahasa Inggris sebagai fokus utama. Audio digunakan untuk mengekstraksi dialog verbal antar karakter yang menjadi objek utama dalam penelitian ini. Proses ekstraksi dilakukan secara digital menggunakan pustaka pemrograman, yang kemudian menghasilkan transkripsi awal dalam bentuk file teks mentah (.txt).

Transkripsi tersebut disimpan sebagai hasil awal untuk keperluan validasi lebih lanjut. Gambar 4.1 menampilkan cuplikan hasil transkripsi dari audio film yang telah dikonversi ke bentuk teks. Namun demikian, karena proses transkripsi otomatis melalui audio dapat mengandung kekeliruan, seperti pemenggalan kalimat yang tidak tepat, hilangnya unsur kata penting, atau gangguan akibat suara latar belakang, maka dilakukan proses verifikasi secara manual. Verifikasi ini dilakukan dengan cara mencocokkan hasil transkripsi dengan subtitle resmi dari masing-masing film untuk menjamin ketepatan isi dialog, susunan kalimat, dan konteks percakapan.

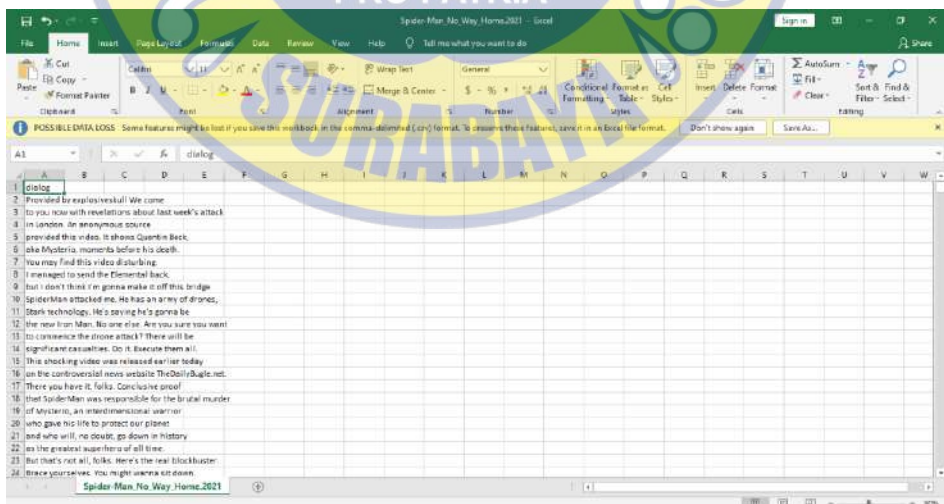
Setelah tahap validasi selesai, data hasil perbaikan disusun ulang agar konsisten dan siap digunakan dalam proses analisis lebih lanjut. Seluruh data dialog

yang telah terverifikasi kemudian diubah ke dalam format CSV guna memudahkan proses pemrosesan berbasis komputasi dan klasifikasi pada tahapan berikutnya. Contoh hasil akhir dari data yang telah disimpan dalam format CSV dapat dilihat pada Gambar 4.2.



Sumber: Dialog film *Spider-Man*, diolah oleh penulis.

Gambar 4. 1 Hasil Transkripsi dari Audio



Sumber: Dialog film *Spider-Man*, diolah oleh penulis.

Gambar 4. 2 Dialog Film

4.2 Hasil Pengolahan Data (*Preprocessing*)

Setelah data dialog dari ketiga film *Spider-Man* berhasil dikumpulkan, tahap berikutnya adalah *preprocessing* untuk mempersiapkan teks sebelum dianalisis dengan algoritma pembelajaran mesin. Proses ini bertujuan menyederhanakan struktur bahasa, menghapus elemen yang tidak relevan, dan menyesuaikan format agar dapat diolah secara komputasional. Hasil *preprocessing* diterapkan pada seluruh dialog yang telah dikonversi ke dalam format .csv dan menunjukkan kesiapan data untuk proses klasifikasi selanjutnya. Tahapan utama *preprocessing* meliputi:

a. *Text Cleaning*

Hasil pembersihan teks menunjukkan banyak dialog mengandung karakter non-alfabet seperti tanda baca, angka, dan simbol. Setelah tahap *cleaning*, elemen tersebut dihapus dan seluruh huruf diubah menjadi kecil, sehingga teks lebih seragam untuk tahap pengolahan berikutnya. Contoh hasil *text cleaning* ditampilkan pada Tabel 4.1.

Tabel 4. 1 *Text Cleaning*

No.	Sebelum proses	Setelah proses
1.	Provided by explosiveskull We come	provided by explosiveskull we come
2.	to you now with revelations about last week's attack	to you now with revelations about last weeks attack
3.	in London. An anonymous source	in london an anonymous source
4.	provided this video. It shows Quentin Beck,	provided this video it shows quentin beck
5.	aka Mysterio, moments before his death.	aka mysterio moments before his death

Sumber: Hasil *preprocessing*, diolah oleh penulis.

b. *Tokenization*

Setelah proses pembersihan teks dilakukan, langkah selanjutnya tokenisasi, yaitu proses memecah kalimat atau frasa menjadi satuan-satuan kata yang disebut token. Tokenisasi berfungsi mengidentifikasi setiap kata secara terpisah, sehingga masing-masing kata dapat dianalisis sebagai fitur individual dalam tahapan pemrosesan selanjutnya. Proses ini menjadi dasar penting dalam analisis teks karena memungkinkan sistem untuk mengenali struktur kalimat secara lebih granular. Hasil dari proses tokenisasi pada data dialog yang telah dibersihkan dapat dilihat pada Tabel 4.2.

Tabel 4. 2 *Tokenization*

No.	Sebelum proses	Setelah proses
1.	provided by explosiveskull we come	['provided', 'by', 'explosiveskull', 'we', 'come']
2.	to you now with revelations about last weeks attack	['to', 'you', 'now', 'with', 'revelations', 'about', 'last', 'weeks', 'attack']
3.	in london an anonymous source	['in', 'london', 'an', 'anonymous', 'source']
4.	provided this video it shows quentin beck	['provided', 'this', 'video', 'it', 'shows', 'quentin', 'beck']
5.	aka mysterio moments before his death	['aka', 'mysterio', 'moments', 'before', 'his', 'death']

Sumber: Hasil *preprocessing*, diolah oleh penulis.

c. *Stopword*

Pada tahap ini, kata-kata umum seperti "*the*", "*without*", dan "*im*" yang tidak memberikan makna signifikan dalam analisis dihapus dari daftar token. Hal ini menghasilkan fokus hanya pada kata-kata bermakna tinggi yang relevan terhadap konteks dialog. Dari contoh sebelumnya, hasil proses *stopword* dapat dilihat pada Tabel 4.3.

Tabel 4. 3 *Stopword*

No.	Sebelum proses	Setelah proses
1.	['provided', 'by', 'explosiveskull', 'we', 'come']	['provided', 'explosiveskull', 'come']
2.	['to', 'you', 'now', 'with', 'revelations', 'about', 'last', 'weeks', 'attack']	['revelations', 'last', 'weeks', 'attack']
3.	['in', 'london', 'an', 'anonymous', 'source']	['london', 'anonymous', 'source']
4.	['provided', 'this', 'video', 'it', 'shows', 'quentin', 'beck']	['provided', 'video', 'shows', 'quentin', 'beck']
5.	['aka', 'mysterio', 'moments', 'before', 'his', 'death']	['aka', 'mysterio', 'moments', 'death']

Sumber: Hasil *preprocessing*, diolah oleh penulis.

d. *Stemming*

Proses stemming mengembalikan kata ke bentuk dasarnya, yang membantu menyatukan variasi bentuk kata yang memiliki makna serupa. Proses ini memperkecil jumlah fitur unik yang akan diproses oleh algoritma klasifikasi. Dari contoh sebelumnya, hasil proses *stemming* dapat dilihat pada Tabel 4.4.

Tabel 4. 4 *Stemming*

No.	Sebelum proses	Setelah proses
1.	['provided', 'explosiveskull', 'come']	['provid', 'explosiveskul', 'come']
2.	['revelations', 'last', 'weeks', 'attack']	['revel', 'last', 'week', 'attack']
3.	['london', 'anonymous', 'source']	['london', 'anonym', 'sourc']
4.	['provided', 'video', 'shows', 'quentin', 'beck']	['provid', 'video', 'show', 'quentin', 'beck']
5.	['aka', 'mysterio', 'moments', 'death']	['aka', 'mysterio', 'moment', 'death']

Sumber: Hasil *preprocessing*, diolah oleh penulis.

4.3 Pemodelan

Pemodelan klasifikasi dalam penelitian ini dilakukan dengan menggunakan algoritma Complement Naive Bayes (CNB) untuk mengelompokkan dialog film ke dalam tiga kategori tema utama, yaitu Konflik, Persahabatan, dan *Multiverse*. Masing-masing kategori dirancang untuk merepresentasikan unsur emosional dan naratif dalam dialog. Pemilihan kata kunci yang tepat dan relevan menjadi faktor penting dalam keberhasilan klasifikasi, mengingat CNB sangat bergantung pada distribusi fitur dalam data. Beberapa kata kunci yang digunakan untuk masing-masing kategori meliputi: “*kill*”, “*enemy*”, dan “*fight*” untuk tema Konflik; “*friend*”, “*help*”, dan “*trust*” untuk tema Persahabatan; serta “*universe*”, “*portal*”, dan “*spell*” untuk tema *Multiverse*. Setelah proses pelatihan model selesai, CNB digunakan untuk memprediksi kategori dari setiap dialog dalam data uji. Hasil klasifikasi menunjukkan bahwa Algoritma Complement Naive Bayes mampu mengelompokkan dialog film ke dalam kategori tematik dengan distribusi hasil yang mencerminkan karakteristik naratif dari masing-masing film.

4.4 Evaluasi

Setelah model Complement Naive Bayes (CNB) selesai dilatih, langkah selanjutnya adalah melakukan evaluasi untuk mengukur performa model dalam mengklasifikasikan dialog ke dalam tiga kategori tematik: Konflik, Persahabatan, dan *Multiverse*. Evaluasi dilakukan menggunakan akurasi klasifikasi serta confusion matrix pada masing-masing film.

4.4.1 *Term Frequency-Inverst Document Frequency*

Dalam penelitian ini, kata-kata penting diekstraksi dari teks dengan mengukur tingkat kekhasan setiap istilah melalui pendekatan Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini mempertimbangkan frekuensi kemunculan suatu kata dalam satu dokumen (TF) serta tingkat kelangkaannya di seluruh dokumen (IDF), sehingga diperoleh bobot yang mencerminkan relevansi kata terhadap konteks film. Secara matematis, perhitungan TF-IDF dirumuskan sesuai dengan persamaan (P4), (P5) dan (P6), dimana t adalah kata, d adalah dokumen, N adalah jumlah total dokumen, dan $DF(t)$ adalah jumlah dokumen yang mengandung kata t . Implementasi dari perhitungan ini dilakukan menggunakan program Python yang dapat dilihat pada Gambar 4.3.

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (P4)$$

Dengan:

$$TF(t, d) = \frac{\text{Jumlah kata } t \text{ dalam } d}{\text{Jumlah semua kata dalam } d} \quad (P5)$$

$$IDF(t) = \log\left(\frac{N}{DF(t)+1}\right) \quad (P6)$$

Pendekatan ini membantu peneliti mengidentifikasi kata-kata paling relevan dalam setiap film, di mana skor TF-IDF tinggi menunjukkan istilah khas yang merepresentasikan tema tertentu. Hasil perhitungan kemudian divisualisasikan dalam grafik batang yang menampilkan distribusi kata dengan skor tertinggi pada masing-masing film, seperti ditunjukkan pada Gambar 4.4, Gambar 4.5, dan Gambar 4.6.

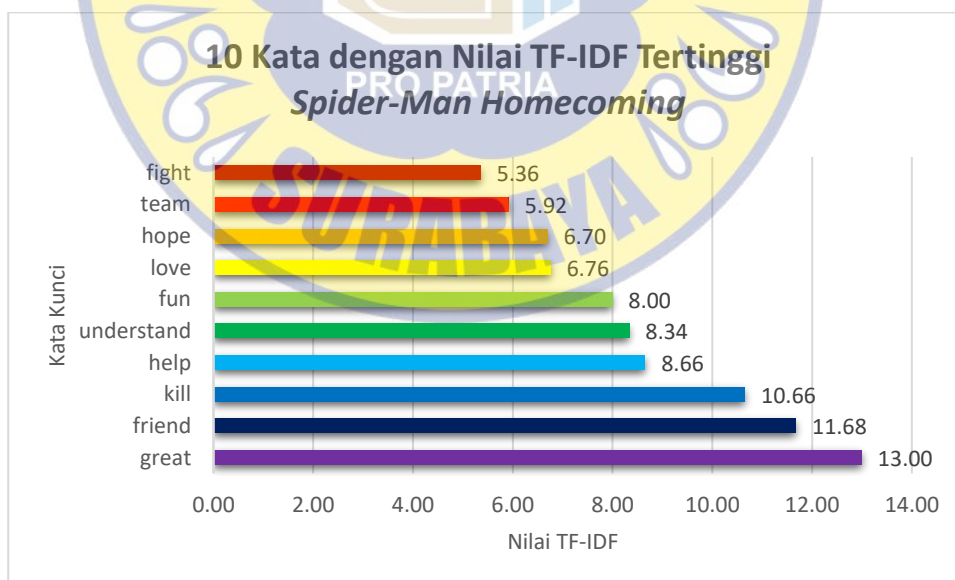
```

27 # === 3. CLEANING DAN FILTERING BERDASARKAN KATA KUNCI ===
28 def clean_and_filter(text):
29     if not isinstance(text, str):
30         return ""
31     text = text.lower()
32     text = re.sub(r'\d+', '', text) # hapus angka
33     text = re.sub(r'^\w\s$', '', text) # hapus tanda baca
34     tokens = text.split()
35     filtered = [word for word in tokens if word in all_keywords]
36     return ' '.join(filtered)
37
38 df['filtered'] = df['stemmed'].apply(clean_and_filter)
39
40 # === 4. TF-IDF VECTORISASI HANYA UNTUK KATA KUNCI ===
41 vectorizer = TfidfVectorizer(vocabulary=sorted(all_keywords)) # hanya kata kunci
42 X_tfidf = vectorizer.fit_transform(df['filtered'])
43
44 tfidf_df = pd.DataFrame(X_tfidf.toarray(), columns=vectorizer.get_feature_names_out())
45
46 # === 5. TOP TF-IDF WORDS ===
47 print("\n10 Kata Kunci dengan Nilai TF-IDF Tertinggi:")
48 top_n = 10
49 tfidf_sums = tfidf_df.sum(axis=0)
50 top_tfidf_words = tfidf_sums.sort_values(ascending=False).head(top_n)
51 print(top_tfidf_words)
52
53 # === 6. BAR CHART 10 KATA KUNCI TERATAS ===
54 plt.figure(figsize=(10, 6))
55 sns.barplot(x=top_tfidf_words.values, y=top_tfidf_words.index, palette='viridis')
56 plt.title("10 Kata Kunci dengan Nilai TF-IDF Tertinggi")
57 plt.xlabel("Nilai TF-IDF")
58 plt.ylabel("Kata Kunci")
59 plt.tight_layout()
60 plt.show()

```

Sumber: Dokumentasi pribadi (hasil coding Python).

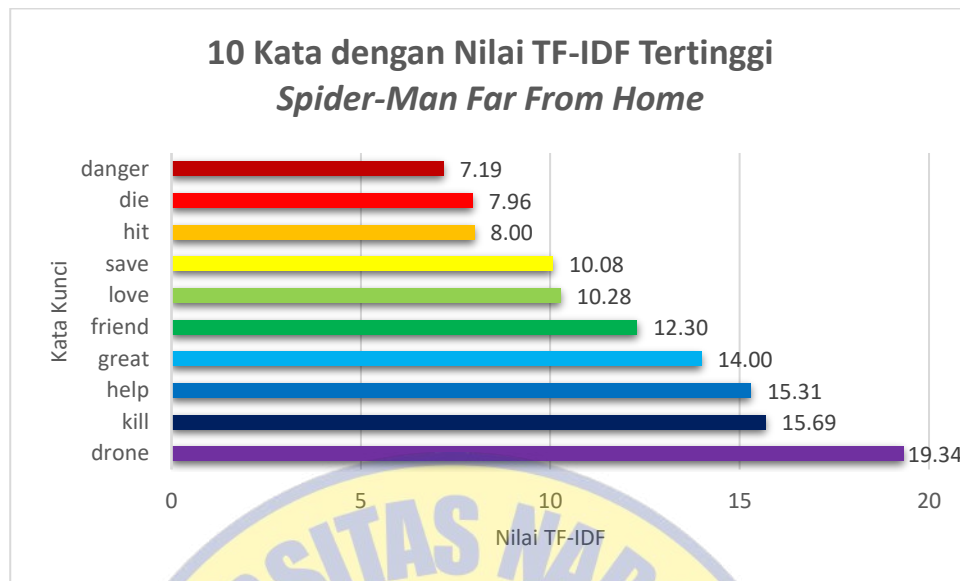
Gambar 4. 3 Source Code TF-IDF



Sumber: Hasil pengolahan TF-IDF menggunakan Python.

Gambar 4. 4 TF-IDF *Spider-Man Homecoming*

Berdasarkan Gambar 4.4 menunjukkan bahwa sepuluh kata kunci dengan nilai tertinggi mencerminkan tema-tema utama yang terkandung dalam dialog karakter. Kata "*great*" memiliki nilai TF-IDF tertinggi sebesar 13.00, diikuti oleh "*friend*" dengan nilai 11.68, yang menunjukkan dominasi nuansa positif dan kedekatan relasi sosial antartokoh, sesuai dengan kategori Persahabatan dalam penelitian ini. Selanjutnya, kata "*kill*" (10.66), "*fight*" (5.36), dan "*help*" (8.66) menggambarkan unsur aksi dan pertentangan, yang berkaitan erat dengan tema Konflik. Sementara itu, kata-kata seperti "*understand*" (8.34), "*hope*" (6.71), dan "*love*" (6.76) mencerminkan kedalaman emosi dan nilai-nilai moral, yang relevan dengan struktur naratif coming-of-age yang menjadi inti karakter Peter Parker. Kata "*fun*" (8.00) dan "*team*" (5.92) juga menunjukkan adanya dinamika kelompok dan keceriaan khas remaja, mendukung konteks sosial dalam narasi. Dengan demikian, analisis TF-IDF ini menegaskan bahwa dialog dalam *Spider-Man: Homecoming* secara eksplisit memuat kosakata yang merepresentasikan tiga kategori tematik utama dalam penelitian, yaitu Konflik, Persahabatan, dan *Multiverse*, sekaligus memperkuat validitas pendekatan klasifikasi berbasis teks yang digunakan. Selain itu, distribusi kata kunci ini juga mengindikasikan bahwa karakter utama memiliki porsi besar dalam membangun nuansa emosional film. Pola kosakata yang muncul memberikan gambaran tentang bagaimana film menyampaikan pesan moral sekaligus konflik internal yang dialami tokoh. Temuan ini juga memperlihatkan konsistensi antara data kuantitatif hasil TF-IDF dengan analisis kualitatif narasi film.

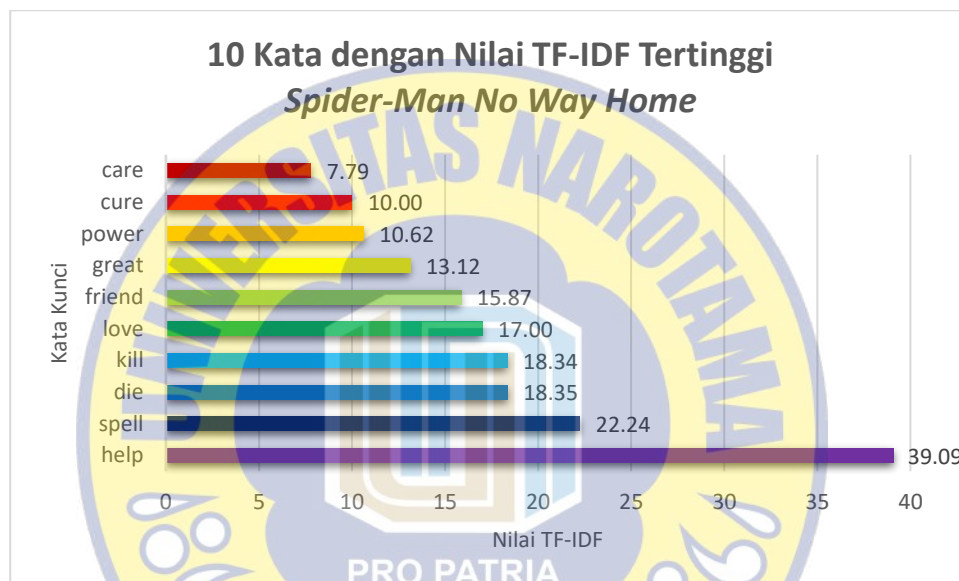


Sumber: Hasil pengolahan TF-IDF menggunakan Python.

Gambar 4. 5 TF-IDF *Spider-Man Far From Home*

Berdasarkan Gambar 4.5 menunjukkan bahwa sepuluh kata kunci dengan bobot tertinggi mengindikasikan tema-tema utama yang muncul secara dominan dalam dialog karakter. Kata "*drone*" memiliki nilai TF-IDF tertinggi sebesar 19,34, menandakan pentingnya elemen teknologi dalam alur cerita film ini. Hal ini sangat relevan dengan tema *Multiverse* dan konflik teknologi yang menjadi inti dari narasi. Selanjutnya, kata "*kill*" (15,69), "*help*" (15,31), dan "*danger*" (7,19) mencerminkan ketegangan serta dinamika konflik yang kuat, sejalan dengan kategori Konflik dalam penelitian. Kata "*great*" (14,00), "*friend*" (12,30), dan "*love*" (10,28) menunjukkan nuansa emosional positif dan hubungan antartokoh, yang menguatkan kehadiran tema Persahabatan. Sedangkan kata "*save*" (10,08), "*hit*" (8,00), dan "*die*" (7,96) juga menekankan situasi genting dan heroik yang sering muncul dalam film, memperkuat keterkaitan dengan konflik moral dan aksi

penyelamatan. Secara keseluruhan, sebaran kata kunci tersebut menunjukkan bahwa dialog dalam film *Spider-Man: Far From Home* didominasi oleh tema-tema seperti konflik yang berkaitan dengan teknologi, hubungan sosial antartokoh, serta tindakan kepahlawanan. Ketiga aspek ini selaras dengan kategori tematik yang digunakan dalam penelitian ini, yaitu Konflik, Persahabatan, dan *Multiverse*.



Sumber: Hasil pengolahan TF-IDF menggunakan Python.

Gambar 4. 6 TF-IDF *Spider-Man No Way Home*

Berdasarkan Gambar 4.6 menunjukkan bahwa sepuluh kata kunci dengan bobot tertinggi mencerminkan dimensi emosional, moral, dan supranatural yang kuat dalam narasi dialog film ini. Kata "*help*" menempati posisi tertinggi dengan nilai TF-IDF sebesar 39,09, menunjukkan dominasi tema bantuan, empati, dan solidaritas antartokoh. Kata "*spell*" (22,24) mengindikasikan peran penting elemen magis dalam alur cerita, yang berkaitan langsung dengan tema *Multiverse* dan dimensi supernatural film. Selanjutnya, kata "*die*" (18,35) dan "*kill*" (18,34)

menguatkan kehadiran unsur tragedi dan konflik emosional yang mendalam. Sedangkan kata "love" (17,00), "friend" (15,87), dan "care" (7,79) menekankan pentingnya hubungan personal dan kasih sayang, yang relevan dengan tema Persahabatan dalam penelitian. Kata "great" (13,12), "power" (10,62), dan "cure" (10,00) mencerminkan perjuangan moral dan tanggung jawab karakter utama terhadap kekuatan dan keselamatan orang lain sebuah ciri khas dari narasi coming-of-age superhero. Dengan demikian, hasil analisis TF-IDF ini mendukung temuan bahwa dialog dalam *Spider-Man: No Way Home* secara eksplisit mengandung kosakata yang mewakili ketiga kategori tematik dalam penelitian, yaitu Konflik, Persahabatan, dan *Multiverse*, serta menunjukkan kompleksitas moral dan emosional yang tinggi dalam struktur ceritanya.

4.4.2 Evaluasi Model: Confusion Matrix

Evaluasi performa model klasifikasi dilakukan menggunakan confusion matrix untuk menilai kemampuan model dalam mengelompokkan dialog film ke dalam tiga kategori tematik, yaitu Konflik, Persahabatan, dan *Multiverse*. Proses evaluasi diawali dengan pemisahan data uji sebesar 20% dari total data secara acak, yang kemudian diproses oleh model Complement Naive Bayes untuk menghasilkan prediksi terhadap kategori setiap baris dialog.

Confusion matrix dalam penelitian ini dibentuk berdasarkan perhitungan menggunakan Persamaan (P3), yang merepresentasikan metode evaluasi performa klasifikasi. Seluruh tahapan evaluasi ini diimplementasikan dalam program Python sebagaimana ditampilkan pada Gambar 4.7, yang mencakup pemisahan data, pelatihan model, prediksi, hingga pembentukan matriks evaluasi. Hasil prediksi

kemudian dibandingkan dengan label aktual untuk menghasilkan confusion matrix, yang memuat jumlah dialog dalam setiap kombinasi label asli dan label prediksi.

```

50 # ----- CNB -----
51 # Encode labels
52 label_encoder = LabelEncoder()
53 df['category_encoded'] = label_encoder.fit_transform(df['category'])
54
55 # Split data
56 X_train, X_test, y_train, y_test = train_test_split(
57     df['stemmed'], df['category_encoded'], test_size=0.2, random_state=42
58 )
59
60 # TF-IDF Vectorization
61 vectorizer = TfidfVectorizer()
62 X_train_tfidf = vectorizer.fit_transform(X_train).toarray()
63 X_test_tfidf = vectorizer.transform(X_test).toarray()
64
65 # Complement Naive Bayes Model
66 cnb = ComplementNB()
67 cnb.fit(X_train_tfidf, y_train)
68
69 # Predict
70 y_pred = cnb.predict(X_test_tfidf)
71
72 # Confusion matrix
73 conf_matrix = confusion_matrix(y_test, y_pred)
74 labels = label_encoder.classes_
75
76 plt.figure(figsize=(8, 6))
77 sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=labels, yticklabels=labels)
78 plt.xlabel("Predicted Label")
79 plt.ylabel("True Label")
80 plt.title("Confusion Matrix Heatmap")
81 plt.show()
82
83 # Perhitungan Akurasi
84 accuracy = accuracy_score(y_test, y_pred)
85 print(f'Accuracy: {accuracy:.2f}')

```

Sumber: Dokumentasi pribadi (hasil coding Python).



Gambar 4.7 Source Code Confusion Matrix

Struktur confusion matrix berbentuk tabel 3x3, dengan sumbu vertikal menunjukkan label sebenarnya (*true label*) dan sumbu horizontal menunjukkan hasil prediksi model (*predicted label*). Karena klasifikasi dilakukan terhadap tiga kelas, hasil evaluasi menghasilkan sembilan kombinasi label-prediksi, yaitu: True Konflik → Konflik (TKK), True *Multiverse* → *Multiverse* (TMM), True Persahabatan → Persahabatan (TPP), Konflik → *Multiverse* (KFM), Konflik → Persahabatan (KFP), *Multiverse* → Konflik (MFK), *Multiverse* → Persahabatan (MFP), Persahabatan → Konflik (PFK), dan Persahabatan → *Multiverse* (PFM).

Tiga kombinasi pertama menunjukkan klasifikasi yang benar (berada pada diagonal utama), sedangkan enam kombinasi lainnya mencerminkan kesalahan prediksi. Penempatan dan besar kecilnya angka pada matriks memberikan informasi rinci tentang kecenderungan model dalam mengenali atau keliru membedakan antar kategori.

Visualisasi hasil evaluasi confusion matrix disajikan dalam bentuk heatmap pada Gambar 4.8, Gambar 4.9, dan Gambar 4.10. Visualisasi ini bertujuan untuk mempermudah identifikasi pola distribusi prediksi model, baik dalam mengenali kategori yang tepat maupun kategori yang sering mengalami kesalahan klasifikasi. Intensitas warna dalam heatmap menunjukkan frekuensi prediksi, semakin gelap warna dalam kotak semakin tinggi jumlah data yang diklasifikasikan pada kombinasi label-prediksi tersebut. Oleh karena itu, kotak berwarna paling gelap menandakan jumlah klasifikasi yang benar paling tinggi, sedangkan warna yang lebih terang menunjukkan prediksi yang lebih sedikit atau kesalahan klasifikasi. Hal ini dapat dilihat pada Tabel 4.5.

Tabel 4. 5 Warna pada Confusion Matrix

Warna	Keterangan
	Menunjukkan prediksi klasifikasi yang benar paling tinggi
	Menunjukkan Prediksi yang lebih sedikit atau kesalahan klasifikasi

Sumber: Hasil penelitian, diolah penulis.

Penjabaran lebih rinci mengenai jumlah prediksi yang benar dan salah disampaikan dalam Tabel 4.6, Tabel 4.7, dan Tabel 4.8. Ketiga tabel tersebut menyajikan detail hasil evaluasi, termasuk label asli, label hasil prediksi, kata kunci dalam dialog, serta jumlah data dalam tiap kombinasi klasifikasi.

Tabel 4. 6 Confusion Matrix *Spider-Man Homecoming*

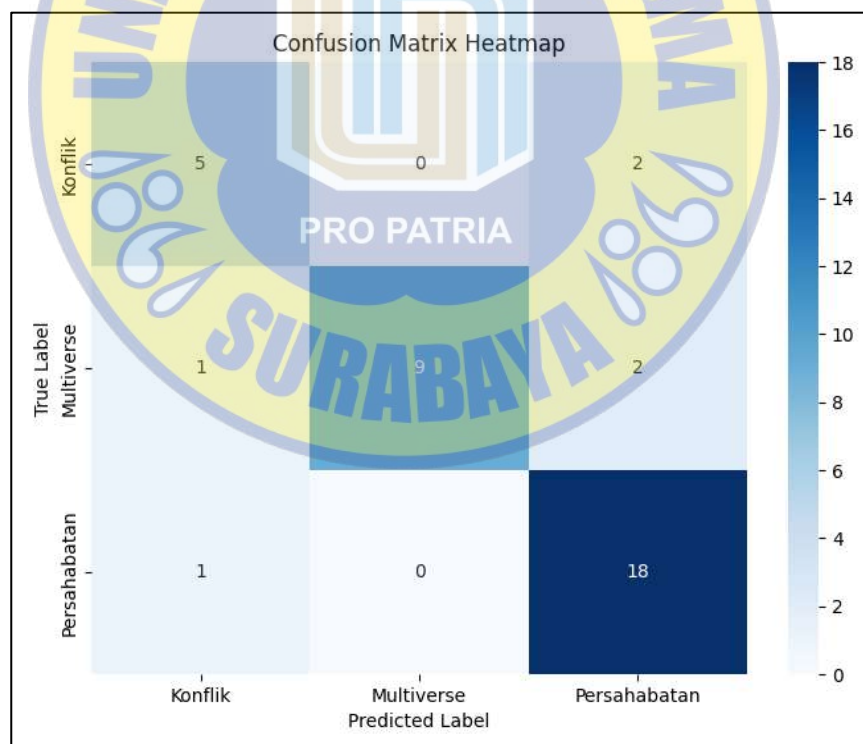
True Label	Prediksi Label	Kata kunci yang ditemukan	Jumlah
Konflik	Konflik	war, death, war, war, danger	5
<i>Multiverse</i>	Konflik	great	1
Persahabatan	Konflik	hug	1
<i>Multiverse</i>	<i>Multiverse</i>	die, great, win, hit, pull, save, hit, win, great	9
Konflik	Persahabatan	war, war	2
<i>Multiverse</i>	Persahabatan	hit, fight	2
Persahabatan	Persahabatan	friend, touch, hope, friend, hug, understand, fun, understand, fun, understand, safe, friend, friend, miss, safe, understand, friend, trust	18

Sumber: Hasil evaluasi klasifikasi.

Berdasarkan Tabel 4.6, model Complement Naïve Bayes menunjukkan kemampuan dalam mengenali kata kunci khas dari setiap kategori tema dalam film *Spider-Man: Homecoming*. Kata-kata seperti war, death, dan danger muncul dominan dalam kategori Konflik, sementara friend, trust, dan hope mendominasi tema Persahabatan. Hal ini mendukung keberhasilan klasifikasi sebanyak 18 data pada kategori Persahabatan dan 5 data pada kategori Konflik. Keberhasilan ini menunjukkan bahwa fitur linguistik dalam tema Persahabatan cenderung lebih konsisten, eksplisit, dan mudah dikenali oleh model. Namun, masih ditemukan kesalahan klasifikasi, seperti dua data bertema Konflik yang diprediksi sebagai Persahabatan, serta dua data Multiverse yang salah diklasifikasikan. Ini mengindikasikan adanya tumpang tindih kosakata dan konteks yang ambigu antartema, terutama dalam dialog yang mengandung dimensi emosional atau reflektif.

Sementara itu, berdasarkan Gambar 4.8, klasifikasi yang tepat berhasil dilakukan pada kategori True Konflik – Prediksi Konflik (TKK) sebanyak 5 data,

True Multiverse – Prediksi Multiverse (TMM) sebanyak 9 data, dan True Persahabatan – Prediksi Persahabatan (TPP) sebanyak 18 data. Di sisi lain, kesalahan klasifikasi muncul pada kombinasi Konflik → Persahabatan (KFP) sebanyak 2 data, Multiverse → Konflik (MFK) sebanyak 1 data, Multiverse → Persahabatan (MFP) sebanyak 2 data, dan Persahabatan → Konflik (PFK) sebanyak 1 data. Temuan ini menegaskan bahwa metode Complement Naïve Bayes paling efektif dalam mengenali tema Persahabatan, yang kemungkinan besar memiliki kosakata yang lebih khas. Sebaliknya, kesalahan klasifikasi pada kategori Konflik dan Multiverse menunjukkan tantangan dalam membedakan konteks yang lebih kompleks atau serupa antar tema.



Sumber: Visualisasi confusion matrix menggunakan Python.

Gambar 4. 8 Confusion Matrix *Spider-Man Homecoming*

Tabel 4. 7 Confusion Matrix *Spider-Man Far From Home*

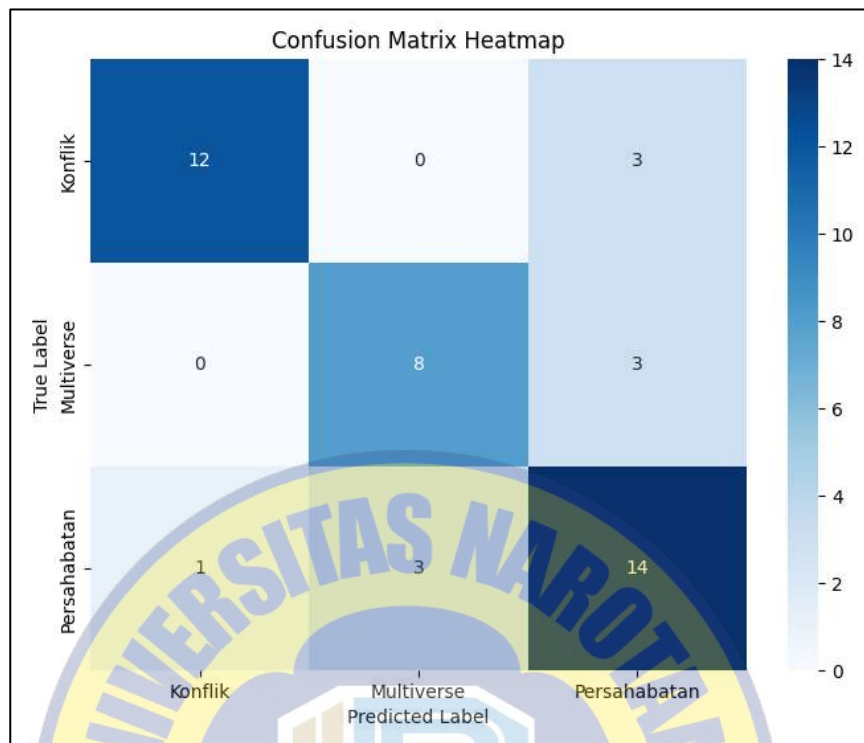
True Label	Prediksi Label	Kata kunci yang ditemukan	Jumlah
Konflik	Konflik	kill, kill, execut, drone, attack, attack, death, war, kill, drone, kill, drone, drone, attack, drone, attack, drone, danger, kill, drone	12
Persahabatan	Konflik	protect	1
<i>Multiverse</i>	<i>Multiverse</i>	fight, die, power, great, hit, save, save, win, die	8
Persahabatan	<i>Multiverse</i>	broke, friend, broke	3
Konflik	Persahabatan	murder, brutal, war, drone, danger	3
<i>Multiverse</i>	Persahabatan	hero, hero, win	3
Persahabatan	Persahabatan	friend, help, help, miss, fun, care, understand, care, love, help, safe, friend, help, trust	14

Sumber: Hasil evaluasi klasifikasi.

Berdasarkan Tabel 4.7 menunjukkan bahwa model Complement Naïve Bayes mampu mengenali kata kunci khas dari setiap kategori tema dalam film *Spider-Man: Far From Home*. Tema Konflik menunjukkan dominasi kata-kata seperti kill, drone, attack dan war yang secara konsisten berhasil dikenali oleh model, menghasilkan klasifikasi tepat sebanyak 12 data. Tema Persahabatan juga memiliki kosakata yang kuat seperti friend, trust, help dan understand, yang membantu model mengklasifikasikan 14 data secara akurat. Keberhasilan ini menunjukkan bahwa model cukup sensitif terhadap fitur-fitur linguistik yang kuat dan khas pada kedua tema tersebut. Namun, masih terdapat beberapa kesalahan klasifikasi, seperti 3 data bertema Konflik yang diprediksi sebagai Persahabatan, dan 3 data bertema Multiverse yang juga diklasifikasikan sebagai Persahabatan. Kesalahan ini mengindikasikan bahwa tema Persahabatan memiliki ekspresi linguistik yang cukup umum atau netral, sehingga berpotensi menimbulkan ambiguitas pada model. Selain itu, tema Multiverse yang sering kali memiliki

konteks yang abstrak atau tidak eksplisit juga cenderung menyulitkan model dalam mengenali ciri khasnya, apalagi jika kata-kata kunci yang muncul mirip dengan konteks emosional atau personal yang dimiliki Persahabatan.

Sementara itu, Gambar 4.9 menampilkan visualisasi confusion matrix dalam bentuk heatmap yang memperkuat pemahaman terhadap performa model. Berdasarkan gambar tersebut, klasifikasi dilakukan dengan benar pada kategori True Konflik - Prediksi Konflik (TKK) sebanyak 12 data, True Multiverse - Prediksi Multiverse (TMM) sebanyak 8 data, dan True Persahabatan - Prediksi Persahabatan (TPP) sebanyak 14 data. Kesalahan klasifikasi muncul pada kategori Konflik - Persahabatan (KFP) sebanyak 3 data, Multiverse - Persahabatan (MFP) sebanyak 3 data, Persahabatan - Konflik (PFK) sebanyak 1 data, dan Persahabatan - Multiverse (PFM) sebanyak 3 data. Ketidaktepatan ini memperlihatkan adanya keterbatasan model dalam membedakan nuansa semantik antar kategori yang bersifat dekat secara emosional maupun naratif. Temuan ini menunjukkan bahwa metode Complement Naïve Bayes cukup efektif dalam mengenali tema Konflik dan Persahabatan, tetapi mengalami tantangan dalam membedakan tema Multiverse, yang kemungkinan memiliki ekspresi atau konteks dialog yang saling tumpang tindih dengan kategori lain, khususnya Persahabatan. Dengan adanya hasil ini, dapat disimpulkan bahwa efektivitas model sangat dipengaruhi oleh kejelasan kosakata tematik yang dimiliki setiap kategori. Oleh karena itu, pendekatan analisis lanjutan dengan metode pembobotan atau representasi teks yang lebih kontekstual dapat menjadi solusi untuk meningkatkan performa klasifikasi pada kategori yang ambigu.



Sumber: Visualisasi confusion matrix menggunakan Python.

Gambar 4. 9 Confusion Matrix *Spider-Man Far From Home*

Tabel 4. 8 Confusion Matrix *Spider-Man No Way Home*

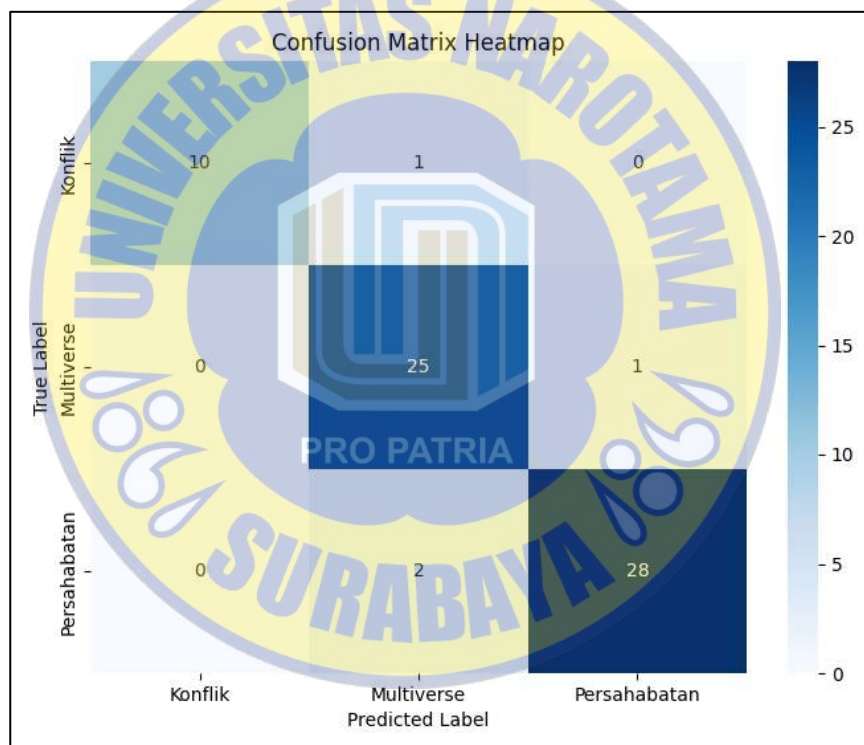
True Label	Prediksi Label	Kata kunci yang ditemukan	Jumlah
Konflik	Konflik	danger, war, kill, murder, brutal, murder, trap, murder, danger, war, attack, drone, attack	10
Konflik	<i>Multiverse</i>	trap	1
<i>Multiverse</i>	<i>Multiverse</i>	spell, power, cure, great, win, swing, spell, spell, win, swing, spell, die, villain, save, save, die, cure, die, spell, portal, hit, great, pull, great, cure, save, power	25
Persahabatan	<i>Multiverse</i>	hope, fun	2
<i>Multiverse</i>	Persahabatan	hit	1
Persahabatan	Persahabatan	broke, love, help, miss, hope, friend, help, safe, help, miss, help, best, friend, heart, help, kiss, help, trust, help, help, help, love, friend, touch, friend, help, best, protect, support	28

Sumber: Hasil evaluasi klasifikasi.

Berdasarkan Tabel 4.8, model Complement Naïve Bayes menunjukkan performa klasifikasi yang sangat baik dalam mengenali tema dialog film *Spider-Man: No Way Home*. Tema Konflik didominasi kata-kata seperti danger, war, kill, murder, dan attack, yang secara konsisten dikenali dengan benar oleh model pada 10 data. Sementara itu, tema *Multiverse* mengandung kata-kata seperti spell, portal, power, dan cure yang menggambarkan konteks dunia alternatif, berhasil diklasifikasikan dengan benar pada 25 data. Tema Persahabatan juga dikenali kuat dengan kehadiran kosakata seperti friend, trust, help, dan support, sehingga menghasilkan klasifikasi yang tepat pada 28 data. Hanya terdapat sedikit kesalahan klasifikasi, seperti 1 data bertema Konflik yang diklasifikasikan sebagai *Multiverse*, 2 data Persahabatan yang salah diklasifikasikan sebagai *Multiverse*, dan 1 data *Multiverse* yang diklasifikasikan sebagai Persahabatan. Hal ini menunjukkan bahwa ekspresi linguistik dalam film ini cenderung lebih eksplisit dan tidak terlalu tumpang tindih antar kategori, sehingga mendukung kinerja klasifikasi model secara keseluruhan.

Gambar 4.10 memperkuat temuan ini melalui visualisasi confusion matrix dalam bentuk heatmap. Klasifikasi dilakukan dengan benar pada kategori True Konflik - Prediksi Konflik (TKK) sebanyak 10 data, True *Multiverse* - Prediksi *Multiverse* (TMM) sebanyak 25 data, dan True Persahabatan - Prediksi Persahabatan (TPP) sebanyak 28 data. Kesalahan klasifikasi yang muncul tergolong sangat kecil, yakni masing-masing hanya 1 data pada kategori Konflik - *Multiverse* (KFM) dan *Multiverse* - Persahabatan (MFP), serta 2 data pada Persahabatan - *Multiverse* (PFM). Tidak ditemukan kesalahan klasifikasi pada kategori Konflik -

Persahabatan (KFP), *Multiverse* - Konflik (MFK), maupun Persahabatan - Konflik (PFK). Temuan ini mengindikasikan bahwa metode Complement Naïve Bayes mampu mengenali dan membedakan karakteristik linguistik dari ketiga kategori tematik secara efektif. Dengan tingkat kesalahan yang sangat rendah, dapat disimpulkan bahwa dialog dalam film *Spider-Man: No Way Home* memiliki struktur bahasa yang lebih terarah dan representatif terhadap tiap tema, sehingga sangat mendukung klasifikasi berbasis tema Konflik, *Multiverse*, dan Persahabatan.



Sumber: Visualisasi confusion matrix menggunakan Python.

Gambar 4. 10 Confusion Matrix *Spider-Man No Way Home*

4.4.3 Akurasi

Pada subbab ini dibahas mengenai perbandingan tingkat akurasi dari model klasifikasi sentimen yang telah diterapkan terhadap tiga film dalam seri *Spider-*

Man, yaitu *Spider-Man: Homecoming*, *Spider-Man: Far From Home*, dan *Spider-Man: No Way Home*. Perbandingan ini dilakukan untuk menilai kemampuan model dalam mengenali serta mengklasifikasikan dialog ke dalam tiga kategori sentimen utama yang telah ditentukan, yaitu Konflik, Persahabatan dan *Multiverse*. Model yang digunakan dalam penelitian ini adalah Complement Naive Bayes, yang dipilih karena kemampuannya dalam menangani data teks dengan banyak fitur. Setiap film dianalisis secara terpisah, dan proses evaluasi dilakukan dengan mengukur akurasi hasil klasifikasi terhadap data uji. Persamaan (P7) digunakan untuk menghitung nilai akurasi, yang implementasinya direalisasikan dalam program Python sebagaimana ditampilkan pada Gambar 4.11. Nilai akurasi diperoleh dengan membandingkan antara label asli dan label hasil prediksi model terhadap data uji.

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \quad (P7)$$

Keterangan:

TP: *True Positive*

TN: *True Negative*

FP: *False Positive*

FN: *False Negative*

Perhitungan akurasi dilakukan secara otomatis dalam program dengan menghitung proporsi data uji yang berhasil diklasifikasikan dengan benar oleh model. Nilai akurasi yang diperoleh dari masing-masing film mencerminkan kemampuan model dalam mengenali pola kata dalam dialog serta menggeneralisasikannya terhadap data baru. Perbedaan akurasi antar film dapat dipengaruhi oleh jumlah data, distribusi kategori, dan karakteristik kosakata yang

husus pada tiap film. Hasil evaluasi ini divisualisasikan pada Gambar 4.12 dalam bentuk perbandingan akurasi model CNB untuk ketiga film, guna mendukung analisis performa secara menyeluruh.

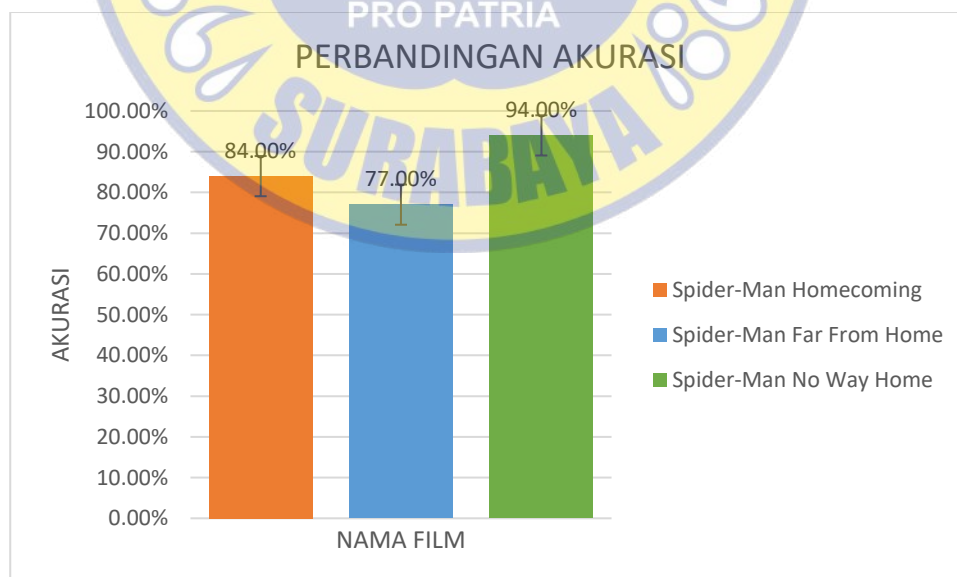
```

72 # Confusion matrix
73 conf_matrix = confusion_matrix(y_test, y_pred)
74 labels = label_encoder.classes_
75
76 plt.figure(figsize=(8, 6))
77 sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xtickl
78 plt.xlabel("Predicted Label")
79 plt.ylabel("True Label")
80 plt.title("Confusion Matrix Heatmap")
81 plt.show()
82
83 # Perhitungan Akurasi
84 accuracy = accuracy_score(y_test, y_pred)
85 print(f'Accuracy: {accuracy:.2f}')
86
87 # Perhitungan Precision
88 precision = precision_score(y_test, y_pred, average='weighted')
89 print(f'Precision: {precision:.2f}')

```

Sumber: Dokumentasi pribadi (hasil coding Python).

Gambar 4. 11 Source Code Akurasi



Sumber: Hasil evaluasi model klasifikasi.

Gambar 4. 12 Perbandingan Akurasi

Berdasarkan grafik tersebut, diperoleh bahwa Film *Spider-Man: No Way Home* memperoleh akurasi tertinggi yaitu 94%, kemudian Film *Spider-Man: Homecoming* berada pada posisi kedua dengan akurasi 84,%, dan Film *Spider-Man: Far From Home* menunjukkan akurasi terendah di antara ketiganya, yaitu 77,%. Perbedaan akurasi ini menunjukkan bahwa tingkat keberhasilan model dalam mengklasifikasikan sentimen dipengaruhi oleh kompleksitas linguistik dan struktur dialog dalam masing-masing film.

Akurasi tertinggi pada *No Way Home* menunjukkan bahwa dialog dalam film ini cenderung lebih eksplisit dan konsisten dalam mengekspresikan ketiga kategori sentimen, sehingga memudahkan proses pembelajaran dan klasifikasi oleh model. Sebaliknya, akurasi terendah pada *Far From Home* kemungkinan disebabkan oleh keragaman gaya bahasa, ekspresi emosional yang ambigu, serta percampuran konteks antar kategori. Kehadiran dialog spekulatif dan humor ringan juga dapat mengaburkan makna sentimen sebenarnya. *Homecoming* menempati posisi tengah dengan akurasi yang cukup stabil, mencerminkan bahwa model masih mampu mengenali pola utama dalam dialog yang berkaitan dengan konflik remaja, hubungan sosial, dan tanggung jawab karakter utama.

4.5 Visualisasi Wordcloud

Pada subbab ini disajikan visualisasi Wordcloud sebagai representasi grafis dari distribusi kata-kata yang paling sering muncul dalam dialog masing-masing film seri *Spider-Man*. Visualisasi ini bertujuan menyajikan gambaran umum terkait kata-kata dominan yang muncul pada setiap film secara terpisah, sehingga

memudahkan peneliti dalam mengidentifikasi istilah-istilah yang merepresentasikan tema utama masing-masing cerita.

Pembuatan Wordcloud dimulai dengan proses pembersihan data dari angka, tanda baca, dan kata-kata umum (*stopwords*). Selanjutnya, kata-kata yang tidak relevan disaring, sehingga hanya tersisa kata-kata yang termasuk dalam daftar kata kunci kategori tematik. Proses ini diimplementasikan melalui program Python sebagaimana ditampilkan pada Gambar 4.13. Secara teknis, Wordcloud dibentuk berdasarkan frekuensi kemunculan kata dalam dialog. Ukuran huruf dalam visualisasi mencerminkan nilai frekuensi tersebut semakin sering kata muncul, semakin besar tampilannya dalam Wordcloud. Hubungan antara frekuensi kata dan ukuran visualisasi dijelaskan secara matematis melalui Persamaan (P8), yang implementasinya juga terintegrasi dalam program pada Gambar 4.13.

$$Ukuran(t) \propto \frac{Frekuensi(t)}{Frekuensi Maksimum} \quad (P8)$$

Dengan demikian, Wordcloud dapat memberikan wawasan visual mengenai distribusi kata-kata penting dalam dialog film, tanpa perlu membaca keseluruhan teks. Teknik ini memungkinkan identifikasi cepat terhadap kata-kata yang paling sering digunakan dan potensial merepresentasikan emosi, konflik, atau tema utama dalam alur cerita. Setiap Wordcloud yang dihasilkan merepresentasikan karakteristik linguistik khas dari masing-masing film dan menjadi penguat dalam interpretasi hasil analisis sentimen berbasis tema. Hasil Visualisasi wordcloud untuk film *Spider-Man* masing-masing ditampilkan pada Gambar 4.14, Gambar 4.15, dan Gambar 4.16. Nilai Frekuensi Wordcloud bisa dilihat pada Tabel 4.9, Tabel 4.10, dan Tabel 4.11.

```

20 # Stopwords dasar (manual)
21 stopwords_manual = set([
22     'yang', 'dan', 'di', 'ke', 'dari', 'untuk', 'pada', 'dengan', 'itu', 'ini', 'karena',
23     'jadi', 'juga', 'ada', 'sudah', 'sangat', 'bisa', 'dalam', 'tidak'
24 ])
25
26 # Gabungkan semua kata kunci
27 all_keywords = set(word.lower() for sublist in categories.values() for word in sublist)
28
29 # Fungsi pembersih teks tanpa NLTK
30 def clean_text(text):
31     if not isinstance(text, str):
32         return ""
33     text = text.lower()
34     text = re.sub(r'\d+', '', text)
35     text = re.sub(r'[^\w\s]', '', text)
36     tokens = text.split()
37     filtered_tokens = [word for word in tokens if word not in stopwords_manual and word in all_keywords]
38     return ' '.join(filtered_tokens)
39
40 # Gabungkan teks dari semua file
41 combined_text = ""
42 for path in file_paths:
43     df = pd.read_csv(path)
44     if 'stemmed' in df.columns:
45         df['filtered'] = df['stemmed'].apply(clean_text)
46         combined_text += ' '.join(df['filtered']) + '\n'
47
48 # Buat dan tampilkan Word Cloud
49 wordcloud = WordCloud(width=800, height=400, background_color='white', colormap='viridis').generate(combined_text)
50 plt.figure(figsize=(12, 6))
51 plt.imshow(wordcloud, interpolation='bilinear')
52 plt.axis("off")
53 plt.title("Wordcloud Spiderman", fontsize=14)
54 plt.show()
55

```

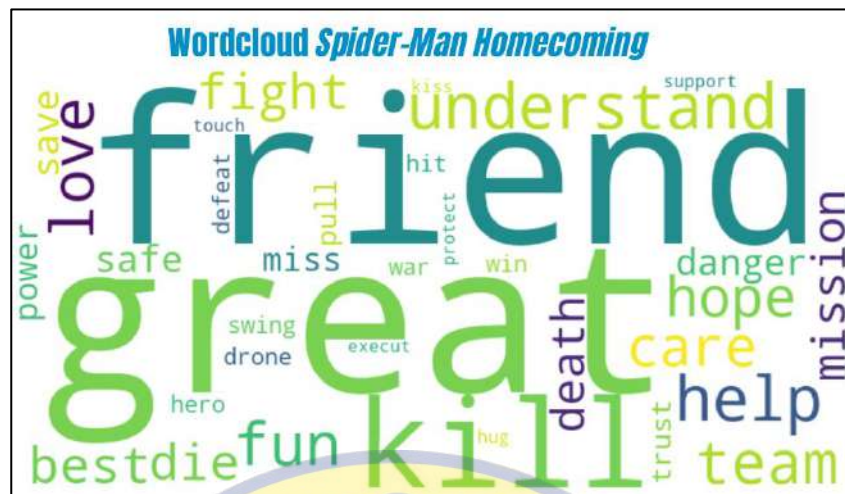
Sumber: Dokumentasi pribadi (hasil coding Python).

Gambar 4. 13 Source Code Visualisasi Wordcloud

Tabel 4. 9 Nilai Frekuensi Wordcloud *Spider-Man Homecoming*

No.	Kata	Nilai	No.	Kata	Nilai	No.	Kata	Nilai
1.	<i>great</i>	14	13.	<i>best</i>	6	25.	<i>defeat</i>	2
2.	<i>friend</i>	13	14.	<i>safe</i>	5	26.	<i>drone</i>	2
3.	<i>kill</i>	12	15.	<i>mission</i>	5	27.	<i>hit</i>	2
4.	<i>understand</i>	9	16.	<i>danger</i>	5	28.	<i>win</i>	2
5.	<i>help</i>	9	17.	<i>safe</i>	5	29.	<i>war</i>	2
6.	<i>fun</i>	8	18.	<i>save</i>	5	30.	<i>swing</i>	2
7.	<i>love</i>	8	19.	<i>death</i>	5	31.	<i>hug</i>	1
8.	<i>team</i>	8	20.	<i>miss</i>	4	32.	<i>execut</i>	1
9.	<i>hope</i>	7	21.	<i>power</i>	4	33.	<i>kiss</i>	1
10.	<i>care</i>	7	22.	<i>trust</i>	3	34.	<i>support</i>	1
11.	<i>fight</i>	6	23.	<i>pull</i>	3	35.	<i>protect</i>	1
12.	<i>die</i>	6	24.	<i>hero</i>	2	36.	<i>touch</i>	1

Sumber: Hasil pengolahan data TF-IDF.



Sumber: Visualisasi wordcloud hasil analisis sentimen.

Gambar 4. 14 Wordcloud *Spider-Man Homecoming*

Berdasarkan Gambar 4.14 dan Tabel 4.9, terlihat bahwa kata-kata seperti "friend", "great", dan "kill" muncul paling menonjol, mencerminkan intensitas penggunaannya dalam dialog karakter sepanjang film *Spider-Man: Homecoming*. Dominasi kata "team" menunjukkan kuatnya unsur kebersamaan, loyalitas, dan dukungan sosial di antara karakter, yang secara langsung mencerminkan tema Persahabatan. Di sisi lain, kemunculan kata seperti "fight" dan "danger" mengindikasikan adanya konflik, baik secara fisik maupun emosional, yang menggambarkan ketegangan serta pertarungan yang dihadapi tokoh utama, sejalan dengan kategori Konflik. Selain itu, kemunculan kata "understand", "hope" dan "help" memperlihatkan dimensi reflektif dan nilai moral yang berkembang dalam diri Peter Parker sebagai karakter utama. Kata-kata ini dapat dikaitkan dengan tema Multiverse, terutama dalam konteks perkembangan karakter, pencarian jati diri, serta dilema etis yang dialaminya. Hal ini menunjukkan bahwa meskipun film

Homecoming belum secara eksplisit mengeksplorasi dunia paralel, elemen-elemen tematik yang mendasari kompleksitas identitas tetap hadir dalam dialog. Dengan demikian, visualisasi ini tidak hanya menampilkan frekuensi kata, tetapi juga memberikan konteks interpretatif yang memperkuat pemahaman terhadap tema dominan serta kaitannya dengan emosi, perkembangan karakter, dan alur cerita film secara menyeluruh. Pola distribusi kata juga membantu menunjukkan keterkaitan antar tema, khususnya antara konflik personal dan nilai moral yang dihadirkan. Dengan begitu, hasil visualisasi berperan penting dalam menghubungkan aspek bahasa dengan dinamika psikologis tokoh utama.

Tabel 4. 10 Nilai Frekuensi Wordcloud *Spider-Man Far From Home*

No.	Kata	Nilai	No.	Kata	Nilai	No.	Kata	Nilai
1.	<i>drone</i>	23	15.	<i>best</i>	6	29.	<i>pull</i>	3
2.	<i>help</i>	21	16.	<i>care</i>	6	30.	<i>win</i>	2
3.	<i>kill</i>	21	17.	<i>hero</i>	5	31.	<i>death</i>	2
4.	<i>friend</i>	17	18.	<i>safe</i>	5	32.	<i>understand</i>	2
5.	<i>great</i>	14	19.	<i>attack</i>	5	33.	<i>support</i>	1
6.	<i>save</i>	12	20.	<i>execut</i>	5	34.	<i>team</i>	1
7.	<i>love</i>	11	21.	<i>touch</i>	4	35.	<i>war</i>	1
8.	<i>trust</i>	9	22.	<i>fun</i>	4	36.	<i>brutal</i>	1
9.	<i>die</i>	9	23.	<i>threat</i>	4	37.	<i>relationship</i>	1
10.	<i>hit</i>	8	24.	<i>murder</i>	4	38.	<i>reveal</i>	1
11.	<i>danger</i>	8	25.	<i>fight</i>	3	39.	<i>fear</i>	1
12.	<i>miss</i>	7	26.	<i>protect</i>	3	40.	<i>broke</i>	1
13.	<i>hope</i>	6	27.	<i>mission</i>	3	41.	<i>dimension</i>	1
14.	<i>power</i>	6	28.	<i>kiss</i>	3	42.	<i>interdimension</i>	1

Sumber: Hasil pengolahan data TF-IDF.



Sumber: Visualisasi wordcloud hasil analisis sentimen.

Gambar 4. 15 Wordcloud *Spider-Man Far From Home*

Berdasarkan Berdasarkan Gambar 4.15 dan Tabel 4.10, terlihat bahwa kata-kata yang paling menonjol dalam dialog adalah "*drone*", "*help*", dan "*kill*", yang muncul dengan ukuran besar, menunjukkan frekuensi tinggi dan mencerminkan fokus cerita pada konflik berbasis teknologi, ancaman fisik, serta upaya penyelamatan. Hal ini sangat relevan dengan tema Konflik dan *Multiverse*, mengingat peran ilusi dan teknologi canggih yang mendominasi narasi film. Kemunculan kata "*friend*", "*care*", dan "*save*" mencerminkan nilai-nilai persahabatan dan kepedulian antar karakter, yang berkaitan erat dengan kategori Persahabatan. Sementara itu, kata "*love*", "*trust*", dan "*hope*" memperkaya dimensi emosional dalam cerita, menegaskan adanya dinamika hubungan dan refleksi personal yang mendalam di antara tokoh-tokohnya.

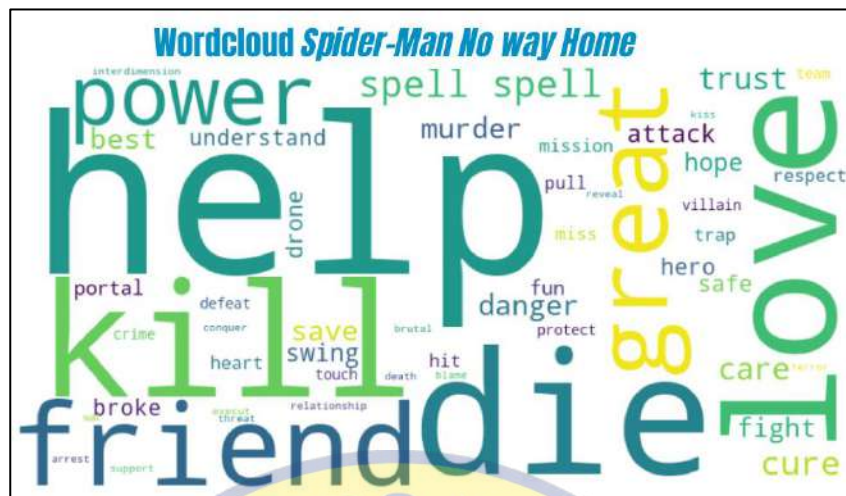
Dengan demikian, analisis frekuensi kata pada Gambar 4.15 menunjukkan bahwa dialog dalam film tidak hanya menyoroiti aspek konflik dan kompleksitas

teknologi, tetapi juga mengangkat nilai-nilai kemanusiaan seperti persahabatan, kasih sayang, dan harapan. Hal ini mengindikasikan bahwa narasi film membangun keseimbangan antara ketegangan aksi dan kedalaman emosional. Visualisasi ini memperkuat pemahaman terhadap konteks tematik dan menjelaskan bagaimana elemen-elemen naratif saling terjalin untuk mendukung tema besar seperti Konflik, *Multiverse*, dan Persahabatan secara terpadu.

Tabel 4. 11 Nilai Frekuensi Wordcloud *Spider-Man No Way Home*

No.	Kata	Nilai	No.	Kata	Nilai	No.	Kata	Nilai
1.	<i>help</i>	40	19.	<i>hope</i>	6	37.	<i>team</i>	2
2.	<i>spell</i>	26	20.	<i>drone</i>	5	38.	<i>crime</i>	2
3.	<i>die</i>	22	21.	<i>hero</i>	5	39.	<i>defeat</i>	2
4.	<i>kill</i>	19	22.	<i>safe</i>	5	40.	<i>death</i>	1
5.	<i>love</i>	18	23.	<i>broke</i>	5	41.	<i>execut</i>	1
6.	<i>friend</i>	18	24.	<i>understand</i>	5	42.	<i>brutal</i>	1
7.	<i>great</i>	14	25.	<i>fun</i>	4	43.	<i>threat</i>	1
8.	<i>power</i>	13	26.	<i>portal</i>	4	44.	<i>war</i>	1
9.	<i>cure</i>	11	27.	<i>hit</i>	3	45.	<i>terror</i>	1
10.	<i>care</i>	9	28.	<i>respect</i>	3	46.	<i>arrest</i>	1
11.	<i>trust</i>	9	29.	<i>miss</i>	3	47.	<i>support</i>	1
12.	<i>murder</i>	8	30.	<i>heart</i>	3	48.	<i>kiss</i>	1
13.	<i>best</i>	8	31.	<i>pull</i>	3	49.	<i>reveal</i>	1
14.	<i>save</i>	8	32.	<i>mission</i>	3	50.	<i>blame</i>	1
15.	<i>danger</i>	8	33.	<i>trap</i>	3	51.	<i>conquer</i>	1
16.	<i>fight</i>	7	34.	<i>protect</i>	2	52.	<i>relationship</i>	1
17.	<i>attack</i>	6	35.	<i>villain</i>	2	53.	<i>interdimension</i>	1
18.	<i>swing</i>	6	36.	<i>touch</i>	2			

Sumber: Hasil pengolahan data TF-IDF.



Sumber: Visualisasi wordcloud hasil analisis sentimen.

Gambar 4. 16 Wordcloud *Spider-Man No Way Home*

Berdasarkan Gambar 4.16 dan Tabel 4.11 menunjukkan bahwa kata-kata kunci yang paling sering muncul dalam dialog karakter. Kata "*help*", "*die*", "*kill*", dan "*love*" tampil paling dominan, mencerminkan tema emosional dan moral yang kuat dalam alur cerita. Kata "*help*" dan "*save*" menyoroti aksi penyelamatan dan kepedulian, yang sejalan dengan tema Persahabatan. Kata seperti "*kill*", "*die*", dan "*danger*" mengindikasikan konflik yang intens, baik secara fisik maupun emosional, yang mencerminkan tema Konflik. Sementara itu, kemunculan kata "*spell*", "*portal*", dan "*multiverse*" mengacu pada aspek supranatural dan perjalanan antardimensi, yang berhubungan langsung dengan tema *Multiverse*. Dengan demikian, frekuensi kemunculan kata-kata kunci dalam dialog karakter pada Gambar 4.10 memperkuat keberadaan tiga tema utama dalam film, yaitu Persahabatan, Konflik, dan *Multiverse*. Dominasi kata-kata seperti "*help*" dan "*save*" menekankan pentingnya solidaritas antar karakter, sementara kata-kata

seperti "*kill*" dan "*danger*" menyoroti ketegangan dan pertentangan dalam alur cerita. Di sisi lain, kehadiran istilah seperti "*spell*", "*portal*", dan "*multiverse*" menegaskan elemen fantasi dan konsep realitas alternatif yang menjadi inti narasi. Hal ini menunjukkan bahwa dialog dalam film berperan signifikan dalam merepresentasikan kompleksitas tema yang diangkat.

Hasil visualisasi Wordcloud tidak hanya menggambarkan kata-kata dominan dalam dialog, tetapi juga dapat menjadi dasar untuk merekomendasikan film berdasarkan tema yang menonjol. Misalnya, film *Spider-Man: Homecoming* menunjukkan kata-kata seperti *friend* dan *trust* yang relevan dengan tema persahabatan. Sedangkan *Spider-Man Far From Home* menampilkan kata *drone* dan *attack* yang mencerminkan konflik dan *Spider-Man No Way Home* memperlihatkan kata *spell* dan *portal* yang mengarah pada tema *multiverse*.

Bagi produser, Wordcloud ini bisa digunakan sebagai bahan evaluasi setelah film dirilis, untuk melihat apakah pesan atau emosi yang ingin disampaikan lewat dialog benar-benar muncul dan terasa sepanjang film. Hal ini bisa dilihat dari kata-kata positif seperti *friend* dan *help* yang konsisten muncul di semua film. Tema-tema tersebut juga sejalan dengan respons umum penonton pada ulasan film yang cenderung menyoroti aspek emosional dan dinamika karakter sebagai bagian yang paling berkesan dari cerita. Bagi penonton sendiri, Wordcloud juga bisa memberi gambaran awal tentang isi dan suasana film. Misalnya, banyaknya kata seperti *kill*, *die*, dan *murder* di *Far From Home* dan *No Way Home* bisa membuat penonton mengira bahwa filmnya akan lebih serius dan penuh konflik, dibandingkan film pertama yang cenderung lebih ringan. Penelitian sebelumnya

oleh Wang Hanmei juga menunjukkan bahwa tema yang muncul dalam naskah film biasanya benar-benar dirasakan oleh penonton, terbukti dari kemiripan yang tinggi antara isi naskah dan komentar mereka. Ini semua memperkuat bahwa Wordcloud bukan hanya sekadar visual, tapi bisa mencerminkan isi emosional film yang nyata dirasakan oleh audiens (Hanmei, 2024).

4.6 Pembahasan

Hasil penelitian ini menunjukkan bahwa klasifikasi sentimen berdasarkan dialog film dapat dilakukan secara efektif dengan menggunakan algoritma Complement Naive Bayes. Algoritma tersebut mampu mengelompokkan teks ke dalam kategori tematik yang merefleksikan kondisi emosional serta alur narasi yang dialami oleh karakter. Dalam konteks penelitian ini, algoritma Complement Naive Bayes berhasil mengklasifikasikan dialog dari tiga film *Spider-Man* ke dalam tiga kategori utama, yaitu Konflik, Persahabatan, dan *Multiverse*. Hasil klasifikasi menunjukkan bahwa model mampu melakukan pengelompokan dengan tingkat akurasi yang cukup tinggi, yakni 84% untuk *Spider-Man: Homecoming*, 77% untuk *Spider-Man: Far From Home*, dan 94% untuk *Spider-Man: No Way Home*. Perbedaan tingkat akurasi ini mencerminkan kompleksitas naratif dan kejelasan ekspresi tematik yang disampaikan dalam dialog masing-masing film.

Proses representasi fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) berhasil mengidentifikasi kata-kata yang menjadi penanda utama dari masing-masing kategori. Kata seperti “fight”, “danger”, dan “enemy” cenderung muncul dalam kategori Konflik, sedangkan “friend”, “trust”, dan “help” muncul pada kategori Persahabatan. Untuk kategori *Multiverse*, kata-

kata seperti “portal”, “dimension”, dan “universe” menjadi penanda dominan. Hasil ini menunjukkan bahwa metode TF-IDF mampu menangkap konteks semantik dari dialog secara efektif dan mendukung klasifikasi tematik yang akurat. Film *No Way Home* menunjukkan akurasi tertinggi, yang dapat dikaitkan dengan struktur narasi yang secara eksplisit menghadirkan tema *multiverse* dan konflik emosional yang mendalam. Dialog dalam film ini mengandung ekspresi yang kuat dan konsisten, sehingga memudahkan model dalam mengenali dan mengklasifikasikan tema. Sebaliknya, film *Far From Home* memiliki akurasi paling rendah, kemungkinan karena gaya dialog yang lebih banyak mengandung ironi, humor, dan ambiguitas tematik. Sementara itu, *Homecoming* berada pada posisi tengah, dengan dialog yang relatif ringan namun tetap mencerminkan distribusi tema yang jelas. Visualisasi wordcloud yang dihasilkan turut mendukung hasil klasifikasi. Setiap kategori menampilkan kumpulan kata dominan yang sesuai dengan karakteristik emosional dan naratif yang dituju. Hal ini memperkuat keandalan pendekatan klasifikasi berbasis fitur kata, karena menunjukkan kesesuaian antara distribusi leksikal dan struktur tema dalam dialog.

Secara teoretis, penelitian ini mendemonstrasikan bahwa dialog film dapat dianalisis menggunakan pendekatan aspect-based sentiment classification untuk memahami struktur emosional dan naratif. Secara praktis, temuan ini memiliki potensi untuk diterapkan dalam pengembangan sistem rekomendasi film, penulisan skenario berbasis emosi, serta pemodelan emosi karakter dalam media fiksi. Selain itu, penggunaan algoritma Complement Naive Bayes terbukti efektif dalam menangani distribusi data yang tidak seimbang, seperti yang umum dijumpai pada

jumlah dialog dalam masing-masing kategori tematik. Dengan demikian, penelitian ini memberikan kontribusi pada pengembangan analisis sentimen berbasis tema, serta membuka peluang eksplorasi lanjutan dalam penerapan teknologi pemrosesan bahasa alami pada media naratif seperti film.

