Klasifikasi Genre Buku Berdasarkan Sinopsis Menggunakan Naïve Bayes dan Logistic Regression

Anak Agung Anom Witaradiani[a1], I Gede Arta Wibawa[a2], Putu Praba Santika[a3]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahua Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]witaradiani.2308561082@student.unud.ac.id
[2]gede.arta@unud.ac.id
[3]praba@unud.ac.id

Abstract

Genre is an important element in book categorization based on specific content characteristics or themes. However, manual classification processes are no longer efficient due to the increasing volume of literature. This study aims to compare the performance of Naïve Bayes and Logistic Regression algorithms in book genre classification based on synopses. The dataset used is secondary data obtained from Kaggle. The dataset consists of 4,535 samples after the preprocessing stage, with feature representation using the TF-IDF method. To address class distribution imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. The evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The experimental results show that Logistic Regression achieved the best performance with 75.19% accuracy and 75.16% F1-score, while Naïve Bayes achieved 72.22% accuracy and 72.11% F1-score. Based on this evaluation, Logistic Regression is considered more effective in classifying book genres from synopsis text.

Keywords: *Book Genre, Text Classification, Naïve Bayes, Logistic Regression, Confusion Matrix*

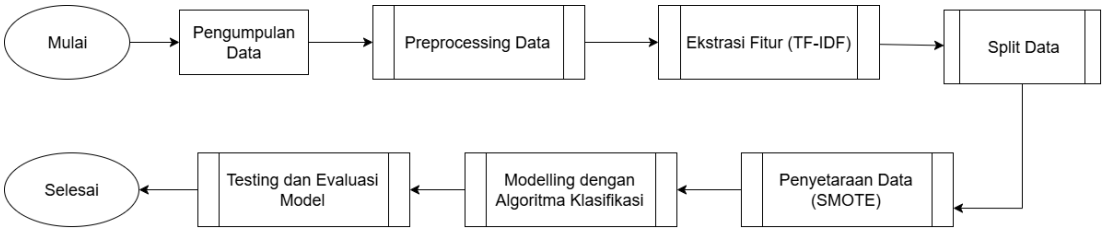
1. Pendahuluan

Genre adalah kategori dengan karakteristik tertentu berdasarkan isi atau tema, baik dalam musik, film, maupun literatur [1]. Dalam buku, genre seperti romance, thriller, fantasy, horror, dan science fiction memudahkan pembaca menemukan bacaan sesuai minat serta membantu pengelola, penerbit, dan platform digital mengatur katalog. Namun, dengan jumlah buku yang terus meningkat pesat setiap tahunnya, baik dalam format fisik maupun digital, proses pengelompokan genre secara manual tidak hanya memakan waktu, tetapi juga berpotensi menimbulkan ketidakkonsistenan klasifikasi akibat subjektivitas manusia [2]. Klasifikasi genre buku dapat dilakukan dengan analisis teks sinopsis karena dianggap mewakili isi cerita. Penelitian sebelumnya menunjukkan Naïve Bayes pada sinopsis novel mencapai akurasi 80,5% dengan 600 data latih dan 200 data uji [3]. Studi lain menggunakan TF-IDF dan Naïve Bayes untuk genre game dengan akurasi 74,65% [1], sedangkan pada data film, Logistic Regression unggul dengan akurasi 58,36% dibanding SVM dan Random Forest [4]. Berdasarkan penelitian sebelumnya, penelitian tentang klasifikasi genre berbasis sinopsis masih jarang yang secara khusus membandingkan kinerja Naïve Bayes dan Logistic Regression pada domain buku. Sebagian besar penelitian sebelumnya hanya fokus pada satu algoritma tertentu atau membandingkan beberapa algoritma pada domain lain seperti film atau *game*. Hal ini menimbulkan kebutuhan untuk mengetahui bagaimana perbedaan performa kedua algoritma tersebut ketika diterapkan pada data sinopsis buku yang memiliki keragaman genre dan ketidakseimbangan distribusi kelas. Oleh karena itu, penelitian ini membandingkan dua algoritma populer yaitu Naïve Bayes dan Logistic Regression untuk melihat sejauh mana perbedaan performa keduanya dalam konteks data genre buku. Pemilihan dua model ini didasari oleh karakteristiknya yang ringan secara komputasi, mudah diimplementasikan, namun memiliki pendekatan yang berbeda. Naïve Bayes mengandalkan probabilitas bersyarat sedangkan Logistic Regression menggunakan pemodelan

linier terhadap peluang. Evaluasi performa dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk menentukan model yang paling optimal dalam konteks klasifikasi genre buku.

2. Metode Penelitian

Penelitian ini mengklasifikasikan genre buku berdasarkan sinopsis melalui tahapan pada Gambar 1. Data sinopsis berlabel dari Kaggle diproses dengan pembersihan teks, stopword removal, tokenisasi, dan stemming. Fitur diekstraksi menggunakan TF-IDF lalu dibagi menjadi data latih dan uji. Untuk mengatasi ketidakseimbangan kelas digunakan SMOTE. Dua model dibangun dengan Naïve Bayes dan Logistic Regression, kemudian dievaluasi menggunakan akurasi, presisi, recall, dan F1-score untuk menentukan model terbaik.



Gambar 1. Alur Tahapan Metode Penelitian

2.1 Pengumpulan Data

Data penelitian ini berupa dataset sekunder dari Kaggle (2022) yang berisi 4.657 buku fiksi berbahasa Inggris dengan label genre tunggal. Setiap entri mencakup sinopsis dan label genre. Secara keseluruhan terdapat sepuluh genre berbeda dengan distribusi jumlah data seperti ditunjukkan berikut.

Tabel 1. Distribusi Jumlah Data Setiap Genre

Genre	Jumlah
Thriller	1.023
Fantasy	876
Science	647
History	600
Horror	600
Crime	500
Romance	111
Psychology	100
Sports	100
Travel	100

2.2 *Preprocessing* Data

Preprocessing merupakan tahap penting untuk menyiapkan data mentah sebelum klasifikasi [5]. Langkah awal mencakup penghapusan baris kosong pada kolom sinopsis atau genre, duplikasi teks, serta sinopsis kurang dari lima kata. Selanjutnya, dilakukan pembersihan dan normalisasi teks melalui tahapan berikut.

a. Data *Cleaning*
Data *cleaning* adalah tahap awal dalam pemrosesan teks yang bertujuan untuk membersihkan dokumen teks dari berbagai jenis *noise* atau gangguan. Data *cleaning*

mencakup proses (mengubah seluruh teks menjadi huruf kecil), penghapusan tautan (URL), angka, tanda baca, serta spasi berlebih [5].

b. Tokenization
Tokenization yaitu memecah kalimat menjadi potongan-potongan kata (token) sebagai satuan analisis [3].

c. *Stopword removal*
Stopword removal yaitu menghapus kata-kata umum seperti “and”, “the”, “is”, dan sebagainya, yang tidak memiliki bobot informasi penting dalam proses klasifikasi [5].

d. Stemming
Stemming yaitu mengubah kata-kata ke bentuk dasar (*root word*) untuk menyederhanakan dan meningkatkan konsistensi representasi kata [5].

2.3 Ekstraksi Fitur dengan *Term Frecuency Inverse Document Frequency* (TF-IDF)

Ekstraksi fitur adalah proses mengubah data teks mentah menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi. Salah satu metode populer dalam ekstraksi fitur teks adalah *Term Frequency–Inverse Document Frequency* (TF-IDF). TF-IDF digunakan untuk merepresentasikan pentingnya suatu kata dalam sebuah dokumen relatif terhadap seluruh korpus dokumen [5]. TF-IDF merupakan kombinasi dari dua komponen yaitu:

a. *Term Frequency* (TF) mengukur seberapa sering sebuah kata muncul dalam satu dokumen [6]. Semakin sering sebuah kata muncul, semakin tinggi nilai TF-nya. Rumus TF sebagai berikut:

$$TF(d,t) = f(d,t) \quad (1)$$

b. *Inverse Document Frequency* (IDF) mengukur seberapa jarang sebuah kata muncul di seluruh dokumen [6]. Kata yang jarang muncul di banyak dokumen akan mendapatkan bobot lebih tinggi. Rumus IDF sebagai berikut:

$$IDF(t) = log\left(\frac{N_d}{df(t)}\right) \quad (2)$$

Sehingga, rumus lengkap TF-IDF untuk term *t* dalam dokumen *d* adalah:

$$TF - IDF = TF(d,t) \times IDF(t) \quad (3)$$

Keterangan:

$f(d,t)$: Frekuensi term t pada dokumen d
N_d : Jumlah dokumen keseluruhan
$df(t)$: Jumlah dokumen yang mengandung term t

2.4 Pembagian dan Penyetaraan Data

Pembagian data bertujuan untuk memisahkan data pelatihan dan pengujian agar model dapat dievaluasi secara objektif. Dalam penelitian ini, data dibagi rasio 80:20, di mana 80% data digunakan untuk melatih model dan 20% untuk menguji model. Jika dilihat kembali pada Tabel 1 terlihat bahwa distribusi jumlah data pada setiap genre tidak seimbang, di mana beberapa kelas seperti sports, travel, romance, dan psychology memiliki jumlah sampel yang jauh lebih sedikit dibandingkan genre mayoritas seperti thriller dan fantasy. Kondisi ini disebut sebagai *imbalanced class* yang dapat menyebabkan model cenderung memprediksi kelas mayoritas sehingga menurunkan performa pada kelas minoritas. Untuk mengatasi masalah tersebut, digunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*). SMOTE yaitu teknik *oversampling* yang menghasilkan data sintetis untuk kelas minoritas berdasarkan interpolasi

antara titik data asli dan tetangga terdekatnya [7]. Secara sederhana, SMOTE menciptakan sampel baru $X\,new$ dengan rumus:

$$X_{new} = X_i + (\hat{X}_t - X_i) \times \delta \quad (4)$$

Keterangan:

X_i : Vektor dari fitur pada kelas minoritas
$\hat{X}_t$: K-nearest neighbors untuk X_i
δ : Angka acak antara 0 sampai 1

2.5 Algoritma Klasifikasi

Klasifikasi merupakan salah satu fungsi dari data *mining* untuk mengelompokkan suatu item data kedalam kategori atau kelas-kelas yang sudah didefinisikan terlebih dahulu [7]. Dalam penelitian ini, dua algoritma klasifikasi digunakan untuk memprediksi genre buku berdasarkan sinopsisnya, yaitu Naïve Bayes dan Logistic Regression.

a. Algoritma Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi probabilistik berbasis Teorema Bayes [3]. Pada klasifikasi teks, ia menghitung probabilitas kelas berdasarkan kemunculan kata. Penelitian ini menggunakan Multinomial Naïve Bayes, yang sesuai untuk data teks dengan representasi frekuensi kata seperti TF-IDF.

Rumus prediksi Naïve Bayes dapat dinyatakan sebagai berikut:

$$P(C_i|\,X) = \frac{P(C_i) \cdot \prod_{j=1}^{n} P\,(x_j|\,C_i)}{P\,(X)} \quad (5)$$

Keterangan:

$P(C_i|\,X)$: Probablitias dokumen X termasuk ke kelas C_i
$P(C_i)$: Probabilitas awal (prior) kelas C_i
$P\,(x_j|\,C_i)$: Probabilitas fitur x_j muncul dalam kelas C_i
$P(X)$: Probabilitas data input (konstanta normalisasi)

b. Algoritma Logistic Regression

Logistic Regression adalah algoritma klasifikasi yang digunakan untuk memprediksi kemungkinan suatu data termasuk ke dalam salah satu kelas [4]. Dalam konteks klasifikasi teks, Logistic Regression memetakan representasi numerik teks (TF-IDF) ke kelas genre, dan memilih kelas dengan probabilitas tertinggi sebagai prediksi akhir.

Rumus prediksi Logistic Regression dapat dinyatakan sebagai:

$$P(y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}} \quad (6)$$

Keterangan:

β_0 : Bias atau intercept
β_n : Koefisien dari masing-masing fitur
x_n : Nilai dari fitur ke-n

2.6 Evaluasi

Tahap evaluasi dilakukan untuk mengukur kinerja dari model klasifikasi yang telah dibangun. Evaluasi dilakukan pada data uji yang dipisahkan dari data latih, menggunakan metrik akurasi,

presisi, recall, dan F1-score yang dihitung dari Confusion Matrix. Confusion Matrix sendiri mengevaluasi kinerja model dengan menyajikan empat komponen utama sebagai dasar perhitungan [2].

- *True Positive* (TP): Data yang benar-benar positif dan diprediksi sebagai positif.
- *True Negative* (TN): Data yang benar-benar negatif dan diprediksi sebagai negatif.
- *False Positive* (FP): Data yang sebenarnya negatif namun diprediksi sebagai positif.
- *False Negative* (FN): Data yang sebenarnya positif namun diprediksi sebagai negative

Dari keempat komponen tersebut, metrik evaluasi dihitung menggunakan rumus berikut:

a. Akurasi (*Accuracy*)
Mengukur proporsi prediksi yang benar terhadap keseluruhan data [2].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

b. Presisi (*Precision*)
Mengukur seberapa akurat prediksi positif yang dibuat oleh model [2].

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

c. Recall
Mengukur seberapa baik model menemukan seluruh data yang sebenarnya positif [2].

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

d. F1-Score
Merupakan rata-rata harmonik dari presisi dan recall, digunakan ketika diperlukan keseimbangan antara keduanya [2].

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

Penggunaan keempat metrik ini penting karena data uji memiliki distribusi kelas yang tidak seimbang. Pada kondisi seperti ini, *accuracy* saja tidak cukup untuk menggambarkan kinerja model secara adil. Misalnya, model dapat memperoleh accuracy tinggi hanya dengan fokus pada kelas mayoritas, namun *precision* dan *recall* dapat mengungkap kelemahan dalam mengenali kelas minoritas. Oleh karena itu, analisis kinerja berdasarkan semua metrik ini memberikan gambaran yang lebih menyeluruh. Selain itu, untuk tujuan perbandingan performa antar model, penelitian ini menggunakan nilai *weighted average* dari *precision*, *recall*, dan *F1-score*. *Weighted average* menghitung rata-rata skor setiap kelas dengan memberikan bobot sesuai jumlah sampel di kelas tersebut, sehingga hasil evaluasi lebih merepresentasikan performa model terhadap distribusi data sebenarnya. Pendekatan ini dipilih agar penilaian kinerja model tetap mempertimbangkan proporsi tiap kelas, meskipun distribusi data tidak seimbang. Rumus perhitungannya sebagai berikut:

$$Weigthed\ Average = \frac{\sum_{i=1}^{n} (m_i \times s_i)}{\sum_{i=1}^{n} s_i} \quad (11)$$

3. Hasil dan Diskusi

Bagian ini menyajikan hasil yang diperoleh dari seluruh tahapan proses penelitian, mulai dari *preprocessing* data teks, ekstraksi fitur menggunakan TF-IDF, penyetaraan data menggunakan metode SMOTE, hingga penerapan dua algoritma klasifikasi, yaitu Multinomial Naïve Bayes dan Logistic Regression.

3.1 Hasil *Preprocessing*

Setelah dilakukan penghapusan terhadap data kosong, duplikat, dan sinopsis yang terlalu pendek, jumlah data yang semula berjumlah 4.657 baris berkurang menjadi 4.535 baris. Data ini kemudian diproses lebih lanjut melalui tahapan *preprocessing* teks. Salah satu contoh dari hasil *preprocessing* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Setiap Tahapan *Preprocessing*

No.	Tahapan	Hasil Setiap Tahapan
1	Data awal	A quiet holiday at a secluded hotel in Devon is all that Hercule Poirot wants, but amongst his fellow guests is a beautiful and vain woman who, seemingly oblivious to her own husband, revels in the attention of another woman's husband. When she is found strangled by powerful hands, were those hands male?
2	Cleaning	a quiet holiday at a secluded hotel in devon is all that hercule poirot wants but amongst his fellow guests is a beautiful and vain woman who seemingly oblivious to her own husband revels in the attention of another womans husband when she is found strangled by powerful hands were those hands male
3	Tokenization	['a', 'quiet', 'holiday', 'at', 'a', 'secluded', 'hotel', 'in', 'devon', 'is', 'all', 'that', 'hercule', 'poirot', 'wants', 'but', 'amongst', 'his', 'fellow', 'guests', 'is', 'a', 'beautiful', 'and', 'vain', 'woman', 'who', 'seemingly', 'oblivious', 'to', 'her', 'own', 'husband', 'revels', 'in', 'the', 'attention', 'of', 'another', 'womans', 'husband', 'when', 'she', 'is', 'found', 'strangled', 'by', 'powerful', 'hands', 'were', 'those', 'hands', 'male']
4	Stopword removal	['quiet', 'holiday', 'secluded', 'hotel', 'devon', 'hercule', 'poirot', 'wants', 'amongst', 'fellow', 'guests', 'beautiful', 'vain', 'woman', 'seemingly', 'oblivious', 'husband', 'revels', 'attention', 'womans', 'husband', 'found', 'strangled', 'powerful', 'hands', 'hands', 'male']
5	Stemming	['quiet', 'holiday', 'seclud', 'hotel', 'devon', 'hercul', 'poirot', 'want', 'amongst', 'fellow', 'guest', 'beauti', 'vain', 'woman', 'seem', 'oblivious', 'husband', 'revel', 'attent', 'woman', 'husband', 'found', 'strangl', 'power', 'hand', 'hand', 'male']

3.2 Hasil Ekstrasi Fitur

Setelah preprocessing, teks diekstraksi menggunakan TF-IDF untuk menghasilkan representasi numerik sinopsis berdasarkan kata atau frasa penting. Setiap baris mewakili dokumen, kolom mewakili fitur, dan nilainya menunjukkan bobot TF-IDF, di mana nilai lebih tinggi berarti kata lebih relevan. Ekstraksi dikonfigurasi agar seimbang antara kelengkapan representasi dan kualitas model. Gambar 2 menampilkan lima dokumen pertama beserta bobot TF-IDF-nya.

	abl	abl escap	abl save	aboard	aboard ship	abus	abus father	...	year	yellow	yfand	young	young child	young princ	zeu
0	0.057121	0.065411	0.00000	0.047625	0.064572	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.076355
2	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.00000	0.000000	0.000000	0.077556	0.068762	...	0.000000	0.000000	0.144342	0.054554	0.073995	0.070228	0.000000
4	0.025402	0.000000	0.05472	0.000000	0.000000	0.000000	0.000000	...	0.028726	0.047757	0.000000	0.000000	0.000000	0.000000	0.000000

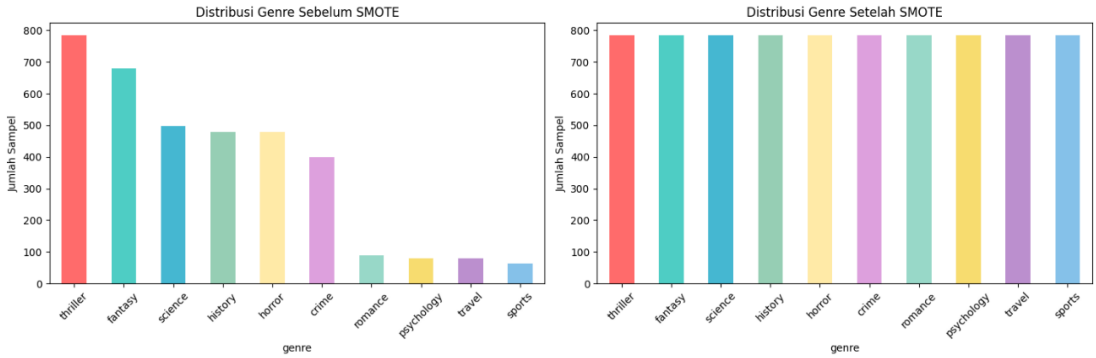
Gambar 2. Hasil Ekstrasi Fitur TF-IDF pada Lima Data Pertama

3.3 Hasil SMOTE

Dari total data, 3.628 digunakan untuk pelatihan dan 907 untuk pengujian. Distribusi genre pada

p-ISSN: 2986-3929
e-ISSN: 3032-1948

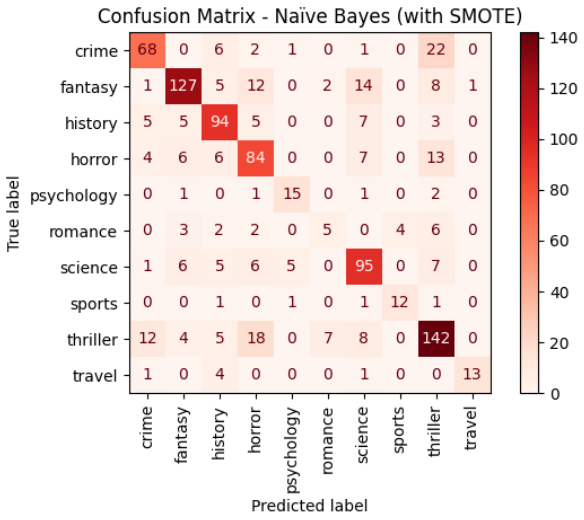
data latih tidak seimbang, dengan kelas minoritas (sports, travel, romance, psychology) jauh lebih sedikit dibanding thriller dan fantasy. Untuk mengatasinya digunakan SMOTE, yang mensintesis sampel baru pada kelas minoritas agar distribusi seimbang. Setelah SMOTE, tiap genre memiliki 784 sampel. Gambar 3 menunjukkan distribusi data sebelum dan sesudah penyetaraan.



Gambar 3. Distribusi Genre Sebelum dan Sesudah SMOTE

3.4 Hasil Klasifikasi dengan Algoritma Naïve Bayes

Setelah pelatihan dengan data latih yang disetarakan menggunakan SMOTE, model Multinomial Naïve Bayes diuji pada 907 sampel data uji. Evaluasi kinerja dilakukan dengan dua metode, yaitu Confusion Matrix dan Classification Report.



Gambar 4. Confusion Matrix pada Algoritma Naïve Bayes

Gambar 4 menunjukkan prediksi Naïve Bayes sebagian besar berada pada diagonal utama confusion matrix. Genre thriller (142), fantasy (127), science (95), crime (68), horror (84), dan history (94) diprediksi dengan akurasi tinggi. Sebaliknya, psychology (15), romance (5), sports (12), dan travel (13) kurang akurat karena sampelnya sedikit. Beberapa genre juga salah diklasifikasikan ke genre lain, seperti crime, horror, romance, fantasy, science, dan thriller.

Tabel 3. Classification Report pada Algoritma Naïve Bayes

Genre	Precision	Recall	F1-Score	Support
Crime	0.7391	0.6800	0.7083	100

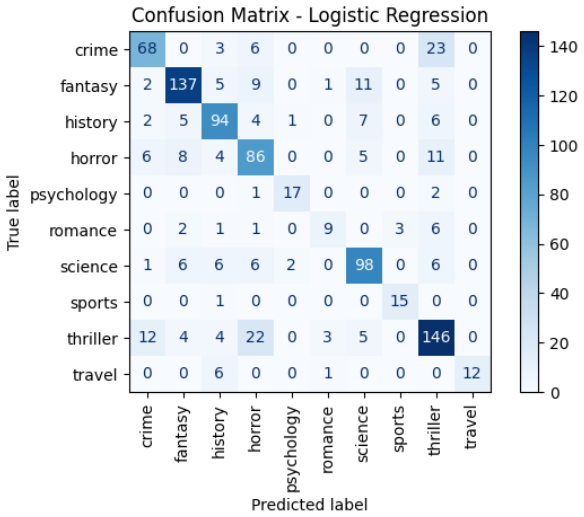
p-ISSN: 2986-3929
e-ISSN: 3032-1948

Genre	Precision	Recall	F1-Score	Support
Fantasy	0.8355	0.7471	0.7611	170
History	0.6462	0.7899	0.6720	119
Horror	0.6462	0.7000	0.6720	120
Psychology	0.6818	0.7500	0.7143	20
Romance	0.3571	0.2273	0.2778	22
Science	0.7037	0.7600	0.7308	125
Sports	0.7500	0.7500	0.7500	16
Thriller	0.6961	0.7245	0.7100	196
Travel	0.9286	0.6842	0.7879	19
Accuracy			0.7222	907
Macro avg	0.7073	0.6813	0.6901	907
Weighted avg	0.7237	0.7222	0.7211	907

Berdasarkan Tabel 3, Naïve Bayes menghasilkan *accuracy* sebesar 0.7222 (72,22%). Nilai *precision* tertinggi terdapat pada genre *travel* (0,9286) dan *fantasy* (0,8355), yang berarti prediksi positif pada genre ini jarang salah. Nilai *recall* tertinggi terdapat pada genre *science* (0,7600), *psychology* (0,7500), dan *sports* (0,7500), menunjukkan kemampuan yang cukup baik dalam menemukan data positif pada kelas tersebut. Nilai *F1-score* tertinggi dimiliki oleh *travel* (0,7879) dan *sports* (0,7500), yang menunjukkan keseimbangan baik antara ketepatan dan kelengkapan prediksi. Sebaliknya, genre *romance* memiliki nilai *precision*, *recall*, dan *F1-score* terendah, yang menunjukkan kesulitan model dalam mengenali genre ini. Secara keseluruhan, nilai *weighted average* pada model Naïve Bayes mencapai *precision* 0,7237, *recall* 0,7222, dan *F1-score* 0,7211.

3.5 Hasil Klasifikasi dengan Algoritma Logistic Regression

Sama seperti pada penerapan algoritma Naïve Bayes, model Logistic Regression juga diuji menggunakan data uji dan dievaluasi melalui dua pendekatan utama, yaitu *Confusion Matrix* dan *Classification Report*.



Gambar 5. Confusion Matrix pada Algoritma Logistic Regression

Dari visualisasi pada Gambar 5, dapat dilihat bahwa sebagian besar prediksi model Logistic Regression berada di diagonal utama *Confusion Matrix*. Genre seperti *thriller* (146), *fantasy* (137), *science* (98), *history* (94), *horror* (86), dan *crime* (68) memiliki tingkat akurasi prediksi yang tinggi dengan nilai diagonal yang dominan. Sementara itu, genre seperti *psychology* (17), *sports* (15), *travel* (12), dan *romance* (9) menunjukkan performa yang lebih rendah karena memiliki jumlah sampel yang sedikit. Namun beberapa genre juga mengalami kesalahan mengklasifikasi ke genre lain.

Tabel 4. Classification Report pada Algoritma Logistic Regression

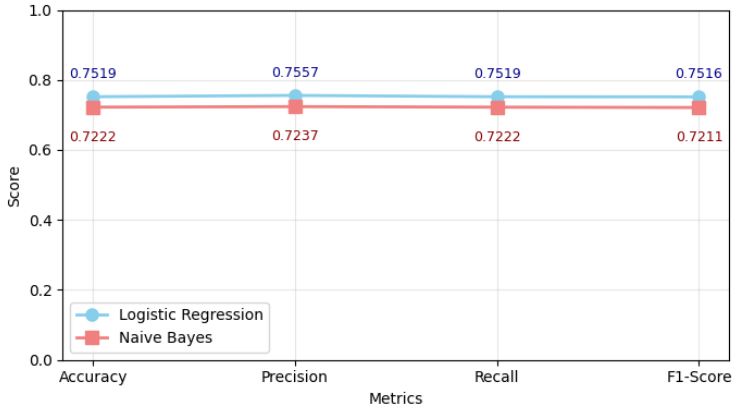
Genre	Precision	Recall	F1-Score	Support
Crime	0.7473	0.6800	0.7120	100
Fantasy	0.8457	0.8059	0.8253	170
History	0.7581	0.7899	0.7737	119
Horror	0.6370	0.7167	0.6745	120
Psychology	0.8500	0.8500	0.8500	20
Romance	0.6429	0.4091	0.5000	22
Science	0.7778	0.7840	0.7809	125
Sports	0.8333	0.9375	0.8824	16
Thriller	0.7122	0.7449	0.7282	196
Travel	1.0000	0.6316	0.7742	19
Accuracy			0.7519	907
Macro avg	0.7804	0.7350	0.7501	907
Weighted avg	0.7557	0.7519	0.7516	907

Berdasarkan Tabel 4, Logistic Regression memperoleh *accuracy* sebesar 0.7519 (75,19%). Nilai *precision* tertinggi dicapai oleh genre *travel* (1,0000), *psychology* (0,8500), dan *fantasy* (0,8457), yang menunjukkan prediksi positif pada genre ini hampir selalu benar. Nilai *recall* tertinggi terdapat pada genre *sports* (0,9375), diikuti oleh *psychology* (0,8500) dan *science* (0,7840), yang berarti model mampu mengenali sebagian besar data positif pada kelas tersebut. Nilai *F1-score* tertinggi dimiliki oleh *sports* (0,8824), *psychology* (0,8500), dan *fantasy* (0,8253), yang mencerminkan keseimbangan optimal antara ketepatan dan kelengkapan prediksi pada genre tersebut. Sebaliknya, genre *horror* memiliki *precision* terendah dan *romance* memiliki *recall* serta *F1-score* terendah. Secara keseluruhan, nilai *weighted average* pada model Logistic Regression mencapai *precision* 0,7518, *recall* 0,7475, dan *F1-score* 0,7486.

3.6 Perbandingan Algoritma Naïve Bayes dan Logistic Regression

Perbandingan kinerja kedua algoritma dilakukan menggunakan nilai *weighted average* dari metrik *precision*, *recall*, *F1-score*, serta nilai *accuracy* yang dihasilkan pada data uji. Pemilihan *weighted average* bertujuan agar hasil evaluasi mencerminkan kinerja model terhadap distribusi data yang tidak seimbang, sehingga kelas dengan jumlah sampel besar maupun kecil tetap diperhitungkan secara proporsional. Berdasarkan Gambar 6, Logistic Regression unggul atas Naïve Bayes di semua metrik evaluasi dengan accuracy 0.7519, precision 0.7557, recall 0.7519, dan F1-score 0.7516. Naïve Bayes hanya mencapai accuracy 0.7222, precision 0.7237, recall 0.7222, dan F1-score 0.7211. Selisih konsisten sekitar 0.03 poin, sehingga Logistic Regression dipilih sebagai model terbaik untuk klasifikasi genre buku berdasarkan sinopsis.

p-ISSN: 2986-3929
e-ISSN: 3032-1948



Gambar 6. Perbandingan Performa Model

4. Kesimpulan

Berdasarkan pada tujuan penelitian dan pengujian yang telah dilakukan, hasil evaluasi menunjukkan bahwa algoritma Logistic Regression memiliki kinerja yang lebih unggul dibandingkan Naïve Bayes, dengan *accuracy* sebesar 0,7475 (74,75%), *weighted average precision* 0,7518 (75,18%), *weighted average recall* 0,7475 (74,75%), dan *weighted average F1-score* 0,7486 (74,86%). Sementara itu, Naïve Bayes mencapai *accuracy* 0,7222 (72,22%), *weighted average precision* 0,7237 (72,37%), *weighted average recall* 0,7222 (72,22%), dan *weighted average F1-score* 0,7211 (72,11%). Logistic Regression lebih unggul pada keseluruhan metrik evaluasi (*accuracy, precision, recall*, dan *F1-score*). Dari hasil tersebut, dapat disimpulkan bahwa Logistic Regression lebih tepat digunakan dalam kasus klasifikasi teks genre buku berdasarkan sinopsis pada dataset ini. Penelitian selanjutnya disarankan untuk mengeksplorasi algoritma lain seperti Support Vector Machine (SVM), Random Forest, maupun pendekatan *deep learning* seperti LSTM atau BERT.

Daftar Pustaka

[1] A. N. Irwan and H. Fahmi, “Classification Game Genre Using TF-IDF and NaÏve Bayes,” *Jurnal Analisis Komputasi Digital*, vol. 9, no. 1, 2025.

[2] F. R. A. Harianto, Z. Alawi, and I. A. Sa’ida, “Pengaruh Komposisi Split Data pada Akurasi Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning,” *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, pp. 36–33, 2025.

[3] V. R. S. Nastiti, S. Basuki, and Hilman, “Klasifikasi Sinopsis Novel Menggunakan Metode Naïve Bayes Classifier,” *Repositor*, vol. 1, pp. 125–130, Dec. 2019.

[4] A. V Siddharth, P. Rakshitha, S. Mazher, V. S. Reddy, and M. G. Uma Maheshwari, “Movie Genre Classification Using Machine Learning,”*International Journal of Innovative Research in Technology*, vol. 11, pp. 4489–4499, May 2025.

[5] Ridwan, E. Heni Hermaliani, and M. Ernawati, “Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian,” *Computer Science (Co-Science)*, vol. 4, pp. 80–88, Jan. 2024, [Online]. Available: http://jurnal.bsi.ac.id/index.php/co-science

[6] A. Sabrani, I. W. Gede Putu Wirarama Wedashwara, and F. Bimantoro, “Metode Multinomial Naive Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia,” *JTIKA*, vol. 2, pp. 89–100, Mar. 2020, [Online]. Available: http://jtika.if.unram.ac.id/index.php/JTIKA/

[7] E. Sutoyo and M. Asri Fadlurrahman, “Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network,” *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 6, pp. 379–385, Dec. 2020.