



## **KLASIFIKASI GENRE ANIME BERDASARKAN SINOPSIS MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS**

**Gede Putra Aditya Brahmantha<sup>1</sup>, Ema Utami<sup>2</sup>, Ainul Yaqin<sup>3</sup>**

<sup>123</sup>Magister Teknik Informatika, Universitas AMIKOM Yogyakarta

Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta

<sup>1</sup>putra.aditya@students.amikom.ac.id, <sup>2</sup>ema.u@amikom.ac.id, <sup>3</sup>ainulyaqin@amikom.ac.id

---

### **Abstract**

*Anime is a popular form of entertainment in the form of animated films originating from Japan, with the popularity of anime, many streaming services provide anime in the content of their services. Anime has a story like a movie, anime also has many genres. In this research, we will classify anime genres based on English synopsis and the algorithm used is K-Nearest Neighbors. Genre classification is done to classify genres into 4 types, namely fantasy, mystery, romance and sports. Each genre contains 100 synopsis data. Preprocessing and TF-IDF weighting are carried out which is continued to the K-Nearest Neighbors classification stage. Based on the evaluation results using 320 training data and 80 test data and the value of  $k = 19$ , the research results obtained are Accuracy of 81.25%, Precision of 83.86%, Recall of 81.25% and F1-Score of 81.17%.*

**Keywords :** *anime, genre, machine learning, k-nearest neighbors, classification*

### **Abstrak**

Anime adalah salah satu bentuk hiburan populer yang berupa film animasi yang berasal dari Jepang, dengan popularitas anime, tidak sedikit streaming services yang menyediakan anime dalam konten layanan mereka. Anime memiliki sebuah cerita layaknya film, anime juga memiliki banyak genre. Dalam penelitian ini akan dilakukan klasifikasi genre anime berdasarkan sinopsis berbahasa Inggris dan algoritma yang digunakan adalah K-Nearest Neighbors. Klasifikasi genre dilakukan untuk mengelompokkan genre menjadi 4 jenis yaitu fantasy, mystery, romance dan sports. Setiap genre berisi 100 data sinopsis. Dilakukan preprocessing dan pembobotan TF-IDF yang dilanjutkan ke tahap klasifikasi K-Nearest Neighbors. Berdasarkan hasil evaluasi menggunakan 320 data latih dan 80 data uji serta nilai  $k=19$ , hasil penelitian yang didapat adalah Akurasi sebesar 81.25%, Precision sebesar 83.86%, Recall sebesar 81.25% dan F1-Score sebesar 81.17%.

**Kata kunci :** *anime, genre, machine learning, k-nearest neighbors, klasifikasi*

---

### **1. PENDAHULUAN**

Anime adalah salah satu jenis animasi yang berasal dari Jepang yang dicirikan oleh ilustrasi yang cerah dan penuh warna yang menggambarkan karakter hidup dalam plot penuh aksi, sering kali dengan tema fantasi atau futuristik. Istilah anime sendiri sering dilekatkan

pada semua produk animasi yang diproduksi di Jepang. [1]

Anime sebagai film animasi juga memiliki banyak Genre, Genre adalah sebuah konsep yang digunakan dalam studi film dan teori film untuk menggambarkan kesamaan antara kelompok-kelompok film berdasarkan estetika atau aspek-aspek sosial, institusional, kultural, dan



psikologis yang lebih luas. Genre film memiliki kesamaan dalam bentuk dan gaya, tema, dan fungsi komunikatif. Dengan demikian, sebuah genre film didasarkan pada seperangkat konvensi yang memengaruhi produksi karya-karya individu dalam genre tersebut dan ekspektasi serta pengalaman penonton. Genre digunakan oleh industri dalam produksi dan pemasaran film, oleh analis dan kritikus film dalam analisis historis film, dan sebagai kerangka kerja bagi penonton dalam pemilihan dan pengalaman film. [2]

Saat ini, pengelompokan genre anime masih dilakukan secara manual oleh publisher dan perlu membaca seluruh sinopsis atau bahkan menonton anime tersebut secara langsung, mengakibatkan investasi waktu yang signifikan dan ditambah pula semakin hari industri anime semakin berkembang pesat dan banyak anime yang terbit dalam waktu bersamaan sehingga, berdasarkan permasalahan ini diperlukan klasifikasi yang efisien seperti penggunaan algoritma K-Nearest Neighbor.

Metode K-Nearest Neighbor (KNN) adalah salah satu teknik klasifikasi dari yang digunakan untuk mengelompokkan data berdasarkan mayoritas kategori. Tujuan utamanya adalah untuk Klasifikasi objek yang baru didasarkan pada atribut dan sampel dari data pelatihan.[3] Salah satu keunggulan K-Nearest Neighbor adalah kemampuannya dalam mengklasifikasikan data yang besar secara efisien dan memberikan hasil yang akurat. [4]

Pada penelitian ini, klasifikasi berita dilakukan dengan mengelompokkan sinopsis ke dalam 4 genre, yaitu fantasy, mystery, romance, dan sports. Jumlah genre berisi 400 data sinopsis. K-Nearest Neighbor digunakan sebagai algoritma dalam penelitian ini dan mengevaluasi hasil klasifikasi dengan confusion matrix serta menghitung recall, precision, dan f1-score.

## 2. TINJAUAN PUSTAKA

### 2.1. Penelitian Terdahulu

Penelitian sebelumnya yang terkait dengan klasifikasi genre film berjudul Klasifikasi Multi-Genre Film berdasarkan Sinopsis menggunakan Naive Bayesian Multi-label (NBML) Classifier yang ditulis oleh Fahry, 2016. Penelitian tersebut bertujuan untuk memberikan solusi pada

masalah penentuan genre film yang memerlukan waktu cukup lama dengan melakukan klasifikasi multi-genre film berdasarkan sinopsis menggunakan naive bayesian multi-label classifier. Dataset yang digunakan adalah teks hasil crawling dari situs imdb. Hasil uji coba menunjukkan bahwa dalam mengklasifikasikan genre film berdasarkan sinopsis, metode paling efektif adalah menggunakan klasifikasi Naive Bayesian Multi-Label (NBML) dengan nilai micro average f1-measure mencapai 87%. Dibandingkan dengan beberapa classifier populer lainnya seperti SVM classifier (65,15%), Decision Tree (50,53%), dan K-NN (59,52%), NBML classifier menunjukkan kinerja yang lebih baik. Penelitian ini menyimpulkan bahwa pendekatan klasifikasi multi-genre film menggunakan naive bayesian multi-label classifier, dengan tahapan-tahapan seperti data crawling data, pre-processing, feature selection, dan klasifikasi yang telah dilakukan, memberikan hasil yang efektif dalam mengklasifikasikan genre film berdasarkan sinopsis.[5]

Selanjutnya Penelitian sebelumnya yang terkait dengan klasifikasi genre film berjudul The Classification of the Movie Genre based on Synopsis of the Indonesian Film yang ditulis oleh Saputra dkk., 2019. Dalam penelitian tersebut data teks sinopsis pada film Indonesia dapat digunakan sebagai fitur untuk menentukan genre film yang sesuai dengan memanfaatkan algoritma machine learning. Algoritma yang digunakan antara lain Naive Bayes, Support Vector Machine, dan Multi-layer perceptron beserta metode ekstraksi fitur yang digunakan yaitu Bag Of Words dan TF-IDF. Jumlah sinopsis dari set data Film Indonesia adalah 1005. Masing-masing genre film memiliki sekitar 200 dataset yang terdiri dari judul film, ringkasan sinopsis, dan genre film. Penelitian ini bertujuan untuk mengeksplorasi beberapa metode ekstraksi fitur dan model pembelajaran mesin untuk mengklasifikasikan genre film. Model yang digunakan menunjukkan bahwa hasil terbaik adalah SVM+TF-IDF Classifier menggunakan unigram dengan akurasi sebesar 0.45%, recall sebesar 45%, precision sebesar 47%, dan skor f1 sebesar 45%. Pada penelitian ini, masih terdapat kekurangan dan kesalahan pada hasil klasifikasi genre film saat melakukan input ringkasan sinopsis. Hal ini disebabkan karena kurangnya dataset terutama melihat jumlah film Indonesia dengan ringkasan sinopsis dan genre film yang masih dapat



digunakan dalam dataset. Perlunya penambahan dataset dan pemilihan data film Indonesia dengan kriteria bahasa Indonesia yang baik dan benar serta menyesuaikan genre film yang ada. [6]

Terdapat pula penelitian yang berkaitan dengan klasifikasi genre yaitu novel, penelitian tersebut berjudul Klasifikasi Sinopsis Novel berdasarkan Jenis Genre menggunakan Multi-class Support Vector Machine dan Chi-square oleh Falakhi, dkk. Penelitian bertujuan mengatasi minimnya informasi genre pada sampul novel dengan mengembangkan sistem klasifikasi berdasarkan sinopsis novel. Algoritma yang digunakan adalah Multi-class Support Vector Machine (SVM) dengan metode One-Against-All. Dalam penelitian tersebut, dilakukan perbandingan antara dua jenis kernel SVM, yakni kernel linear dan kernel gaussian, dengan tujuan memperoleh tingkat akurasi klasifikasi yang optimal. Pendekatan TF-IDF dan Chi-square digunakan untuk memberikan bobot dan menyeleksi fitur dari sinopsis novel. Terdapat 240 teks sinopsis yang dibagi menjadi 4 jenis genre, yaitu horror, romantis, sci-fi, dan sejarah, digunakan sebagai dataset pelatihan dan pengujian. Hasil uji yang dilakukan menunjukkan bahwa ketika menggunakan kernel linear pada SVM dengan threshold Chi-square sebesar 80%, parameter sequential training  $\lambda = 0,5$ ,  $\gamma = 0,05$ , complexity (C) = 1, epsilon ( $\epsilon$ ) = 0,0001, dan jumlah iterasi maksimum adalah 100 iterasi akurasi klasifikasi mencapai puncaknya pada 94,58%. Sistem klasifikasi yang dikembangkan tersebut mampu memberikan akurasi klasifikasi tinggi (94,58%) dalam mengidentifikasi genre novel berdasarkan informasi sinopsis. [7]

Penelitian lainnya adalah klasifikasi genre film yang berjudul On the Use of Synopsis-based Features for Film Genre Classification yang ditulis oleh Portolese dan Feltim, 2018. Dalam penelitian tersebut mengklasifikasikan sinopsis berbahasa Portugis Brazil ke dalam setidaknya salah satu dari 12 label genre yang ada, eksperimen tersebut menggunakan 19 metode ekstraksi fitur yang dikombinasikan dengan 4 algoritma klasifikasi multi-label yaitu Decision Tree Classifier, Random Forest Classifier, Extra Trees Classifier, dan multi-layer perceptrons. Sinopsis diekstraksi pada bulan Februari 2018 menggunakan API TMDb yang tersedia. Hanya sinopsis dalam bahasa Portugis Brasil yang dipilih, dengan total 13.449 sinopsis. Ekstraksi

fitur dilakukan menggunakan TF-IDF dan Word Embedding model Doc2Vec. Semua kombinasi dievaluasi menggunakan 5-fold cross validation dan metrik klasifikasi akhir diperoleh dengan merata-rata hasil dari setiap lipatan. Dengan menghitung rata-rata antara nilai-nilai di semua genre, hasil yang diperoleh dari percobaan terbaik menggunakan ekstraksi fitur TF-IDF dan algoritma Multi-Layer Perception menghasilkan precision 57,61%, recall 53,36%, dan f1-score sebesar 54,80%. Diperhatikan bahwa tidak ada teknik ekstraksi fitur berbasis embedding yang dapat mengungguli pendekatan tf-idf. Penelitian tersebut memberikan pandangan positif pada masalah klasifikasi genre film, karena f1-skor 54,8% yang diperoleh dalam percobaan terbaik dapat ditingkatkan lebih lanjut dalam penelitian di masa depan di mana fitur dan pengklasifikasi yang lebih spesifik dapat digunakan. [8]

## 2.2. Anime

Anime merupakan bentuk hiburan populer dari Jepang yang memiliki cakupan luas di kalangan penonton karena menyajikan berbagai genre yang dapat dinikmati oleh berbagai kelompok usia. Sebagian besar anime yang diproduksi merupakan adaptasi dari manga, yakni komik Jepang. Keunggulan anime sebagai medium hiburan massa yang bersifat audio visual adalah kemampuannya untuk mentransfer budaya kepada penontonnya. Dengan berbagai genre yang ditawarkan, masyarakat memiliki banyak pilihan dalam memilih jenis anime yang sesuai dengan preferensi mereka. Terus bermunculannya berbagai judul anime mencerminkan respons terhadap permintaan pasar, dengan setiap judul berusaha menyajikan cerita yang beragam dan unik sesuai dengan segmentasi pasar yang dituju. [9]

Anime dan animasi barat memiliki karakteristiknya masing-masing. Perbedaan yang sangat jelas terlihat pada unsur artistik keduanya, baik secara artistik maupun cerita. Animasi barat memiliki style atau gaya gambar karakter yang lebih menyerupai bentuk karakter manusia asli mulai dari bentuk tubuh, wajah, dan rambut. Dari segi pembuatannya pun tidak berasal dari komik melainkan dari storyboard atau alur cerita yang dibuat oleh penulis naskah yang kemudian diarahkan oleh sutradara dan biasa disebut dengan Original Works. Cakupan



genrenya juga sangat terbatas seperti komedi, romance, super power, dan action, sedangkan anime memiliki style atau gaya gambar yang khas Jepang seperti bentuk mata yang besar, dagu yang lancip, dan gaya rambut karakter yang memiliki berbagai macam model serta yang terakhir adalah desain wajah yang tampan dan cantik. Pembuatan sebuah anime biasanya diambil atau diadaptasi dari manga yang sedang tren atau naik daun, namun ada juga beberapa anime yang dibuat sebagai karya orisinal. Berbeda dengan Animasi Barat, genre Anime memiliki cakupan yang luas seperti percintaan, kehidupan sehari-hari, petualangan, sihir, kekuatan super, supranatural, misteri, aksi, dll.[10]

### 2.3. Crawling

Web Crawler adalah sebuah program/perangkat lunak atau skrip terprogram yang menjelajahi World Wide Web dengan cara yang sistematis dan otomatis. Struktur WWW adalah struktur grafis, yaitu tautan yang disajikan dalam sebuah halaman web dapat digunakan untuk membuka halaman web lainnya. Internet adalah sebuah graf berarah dimana halaman web sebagai simpul dan hyperlink sebagai sisi, sehingga operasi pencarian dapat diringkas sebagai proses menelusuri graf berarah. Dengan mengikuti struktur terhubung dari Web, web crawler dapat melintasi beberapa halaman web baru yang dimulai dari sebuah halaman web. Web crawler bergerak dari satu halaman ke halaman lain dengan menggunakan struktur grafis dari halaman web. Program seperti ini juga dikenal sebagai robot, spider, dan worm. Perayap web dirancang untuk mengambil halaman web dan memasukkannya ke dalam repositori lokal. Crawler pada dasarnya digunakan untuk membuat replika dari semua halaman yang dikunjungi yang kemudian diproses oleh mesin pencari yang akan mengindeks halaman yang diunduh yang membantu dalam pencarian cepat. Tugas mesin pencari adalah menyimpan informasi tentang beberapa halaman web, yang mereka ambil dari WWW. Halaman-halaman ini Halaman-halaman ini diambil oleh Web crawler yang merupakan peramban Web yang otomatis mengikuti setiap tautan yang dilihatnya. [11]

### 2.4. K-Nearest Neighbors

K-Nearest Neighbors (KNN) adalah suatu algoritma yang digunakan untuk mengidentifikasi tetangga terdekat. KNN adalah algoritma yang sederhana, yang beroperasi dengan mencari jarak terpendek antara data pengujian dan data pelatihan untuk menemukan K-tetangga terdekat (KNN), lalu mengambil mayoritas dari KNN ini untuk memprediksi data pengujian. Inti dari KNN adalah mencari K objek dari data pelatihan yang memiliki jarak paling dekat dengan data pengujian.[12]

### 2.5. Confusion Matrix

Prinsip perhitungan akurasi data mining dan text mining biasanya dilakukan dengan menggunakan pendekatan confusion matrix. Recall, akurasi, dan f1-score adalah tiga output dari perhitungan yang dilakukan oleh rumus ini. [13]

1. Recall adalah rasio kasus positif yang terdeteksi dengan tepat.

$$Recall = \frac{TP}{(TP+FN)} \quad (1)$$

2. Precision adalah rasio kasus positif yang diidentifikasi dengan tepat.

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

3. Akurasi adalah perbandingan kasus yang diidentifikasi dengan benar terhadap total kasus..

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (3)$$

4. F1-Score adalah nilai rata-rata harmonik dari recall dan precision.

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (4)$$

Confusion Matrix atau matriks kebingungan adalah tabel yang sering digunakan untuk menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya telah diketahui. Tabel berikut menunjukkan Confusion Matrix untuk model klasifikasi dua kelas.[14]:

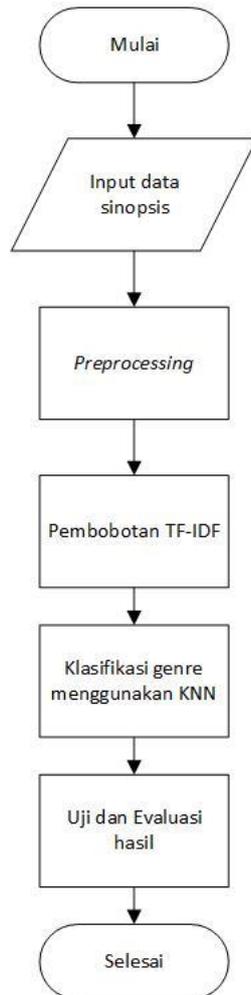
**Tabel 1.** Contoh Confusion Matrix

	Prediksi 0	Prediksi 1
Aktual 0	TN (True Negative)	FP (False Positive)
Aktual 1	FN (False Negative)	TP (True Positive)

### 3. METODOLOGI PENELITIAN

#### 3.1. Alur Penelitian

Pada gambar 1 menunjukkan alur penelitian yang dilakukan.



**Gambar 1.** Alur Penelitian

#### 3.2. Pengumpulan Data

Pada penelitian ini, data diperoleh dengan melakukan web crawling untuk mendapatkan teks sinopsis dengan genre 'Fantasy' 'Mystery' 'Romance' dan 'Sports' dari situs Anime-Manga Database berbahasa Inggris sebanyak 100 sinopsis untuk masing-masing genre sehingga total data yang diperoleh adalah 400 artikel. Data tersebut diberi label sesuai dengan genre anime. Dari dataset yang digunakan, data dibagi dengan proporsi 80% untuk data training dan 20% untuk data testing, proses pembagian data dilakukan secara acak.

#### 3.3. Preprocessing

Untuk mengatasi ketidakteraturan dalam struktur data teks, langkah awal yang diperlukan adalah menyiapkan teks tersebut agar dapat diubah menjadi data yang terstruktur. [15] Preprocessing data sering diabaikan dalam kegiatan penambangan data, yang dapat mengakibatkan kurangnya kendali terhadap data dan hasil analisis yang kurang akurat [16]. Tahap preprocessing pada penelitian ini terdiri dari case folding, cleansing, stopword removal, stemming, dan tokenization seperti yang terlihat pada Gambar 2.



**Gambar 2.** Alur Preprocessing

##### 3.3.1. Case Folding

Case folding adalah langkah pertama dalam tahap prapemrosesan yang bertujuan agar setiap huruf berubah menjadi huruf kecil.

##### 3.3.2. Cleansing

Cleansing adalah langkah menghilangkan kata-kata yang tidak relevan dari dokumen guna



mengurangi noise seperti nama pengguna, tanda baca, dan URL.[17] Selain itu, angka juga akan dihilangkan pada tahap ini.

### 3.3.3. Stopword Removal

Stopword merujuk pada kumpulan kata-kata umum yang dianggap tidak esensial dan tidak relevan dalam proses klasifikasi. Tahap eliminasi Stopword dilakukan dengan menghilangkan kata-kata yang umumnya ditemui namun tidak memiliki dampak signifikan terhadap makna suatu teks atau kalimat.[18]

### 3.3.4. Stemming

Proses stemming adalah salah satu proses awal yang penting, tidak hanya untuk mengurangi ukuran database dan meningkatkan efektifitas, tetapi juga karena proses stemming mengembalikan kata kata ke kata dasarnya yang mengandung arti dari kata tersebut [19]. Dalam penelitian ini, akan digunakan library Porter Stemmer untuk melakukan proses stemming pada teks sinopsis.

### 3.3.5 Tokenization

Tokenisasi bekerja dengan cara memisahkan kata-kata dari teks atau kalimat menjadi beberapa token. Proses ini tidak akan menyertakan spasi.

## 3.4. TF-IDF

TF-IDF telah digunakan secara luas di bidang text mining untuk mengevaluasi hubungan setiap kata dalam koleksi dokumen. Secara khusus, TF-IDF digunakan untuk mengekstraksi kata-kata inti (yaitu, kata kunci) dari dokumen, menghitung derajat yang sama di antara dokumen, menentukan peringkat pencarian, dan sebagainya. TF dalam TF-IDF berarti kemunculan kata-kata tertentu dalam dokumen. Kata-kata yang memiliki frekuensi term (TF) tinggi memiliki relevansi yang signifikan dalam dokumen tersebut. Sebaliknya, Document Frequency (DF) mengindikasikan seberapa sering kata tertentu muncul dalam seluruh koleksi dokumen. DF menghitung kemunculan kata tersebut dalam berbagai dokumen, bukan hanya dalam satu. Kata-kata yang memiliki nilai DF tinggi kurang signifikan karena cenderung muncul dalam sebagian besar dokumen. Oleh karena itu, IDF yang merupakan kebalikan dari

DF digunakan untuk mengukur tingkat kepentingan kata di semua dokumen. Nilai IDF yang tinggi berarti kata-kata yang jarang muncul di semua dokumen, yang mengakibatkan peningkatan nilai penting. [20]

Rumus umum untuk Tf-Idf adalah [21]:

$$Wdt = TFdt * IDFt \quad (5)$$

Keterangan :

d = dokumen ke-d

t = kata ke-t dari kata kunci

W = bobot dokumen ke-d terhadap kata ke-t

tf = frekuensi kemunculan kata kunci pada suatu dokumen

IDF = Inverse Dokument Frequency

D= total dokumen

df = jumlah dokumen yang mengandung kata kunci yang dicari

## 3.5. Implementasi K-Nearest Neighbor

Tahapan dalam algoritma KNN dalam melakukan klasifikasi adalah sebagai berikut [22] :

1. Tentukan jumlah k yang digunakan
2. Melakukan Vektorisasi dari dataset yang digunakan
3. Menghitung jarak testing ke seluruh data training yang ditentukan dengan perhitungan euclidean distance

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (6)$$

Keterangan :

D = Jarak

x = Data Train

y = Data Testing

4. Ambil tetangga sebanyak nilai k yang paling dekat dengan data testing berdasarkan perhitungan jarak sebelumnya
5. Didapatkan hasil prediksi berdasarkan anggota tetangga terbanyak suatu label

## 3.6. Evaluasi Hasil

Pengujian dilakukan dengan menggunakan confusion matrix untuk mengetahui nilai dari



precision (2), recall (1), f-measure (4), dan akurasi(3). Perhitungan precision, recall, f-measure, dan akurasi dilakukan di setiap pengujian dan dicari nilai total rata-rata dari setiap kelas.

**4. HASIL DAN PEMBAHASAN**

Penelitian ini menggunakan data yang didapatkan dengan melakukan web crawling untuk mendapatkan sinopsis dari situs anime-manga database berbahasa inggris, terdapat 400 data yang dikumpulkan dari web crawling yang dibagi menjadi 100 genre fantasy, 100 genre mystery, 100 genre romance, dan 100 genre sports. Dari dataset yang digunakan, data dibagi menjadi proporsi 80% untuk data training dan 20% untuk data testing, proses pembagian data dilakukan secara acak dan setiap label mempunyai anggota yang sama saat dilakukan pembagian.

Data harus melalui tahap preprocessing, yaitu mengubah semua huruf menjadi huruf kecil, lalu menghapus teks yang tidak relevan, menghapus kata-kata umum (stopwords), menemukan kata-kata ke bentuk dasarnya, dan pada tahap terakhir memisahkan semua kalimat menjadi kata-kata yang terpisah (tokenisasi). Setelah tahap preprocessing, pembobotan TF-IDF dilakukan dengan menggunakan rumus (5), setelah bobot didapatkan, klasifikasi dilakukan dengan menggunakan algoritma K-Nearest Neighbors (KNN). Nilai k yang digunakan dalam penelitian ini mengikuti aturan umum yang ada yaitu akar dari jumlah data training yang menghasilkan k=19.

Untuk mengevaluasi sistem, dilakukan perhitungan total data yang diklasifikasikan dengan benar dibagi oleh jumlah keseluruhan data uji. Informasi mengenai hasil klasifikasi ditampilkan dalam Tabel 2, yaitu confusion matrix yang memberikan detail prediksi sistem untuk setiap kelas.

**Tabel 1.** Hasil Confusion Matrix

	Fantasy	Mystery	Romance	Sports
PREDIKSI	(A)	(B)	(C)	(D)
AKTUAL				
Fantasy (A)	18	2	0	0
Mystery (B)	4	14	1	1
Romance (C)	6	1	13	0
Sports (D)	0	0	0	20

Evaluasi dilaksanakan untuk menilai apakah penelitian telah mencapai tujuannya atau belum. Evaluasi penelitian ini melibatkan perhitungan akurasi, presisi, recall, dan F1-Score berdasarkan hasil klasifikasi yang terdokumentasikan dalam confusion matrix. Berikut hasil perhitungan manual dalam eksperimen ini berdasarkan confusion matrix yang telah diperoleh

**Perhitungan Akurasi**

$$Accuracy = \frac{(Total\ Data\ Benar)}{(Total\ Dataset\ Testing)}$$

$$= \frac{(18+14+13+20)}{(80)} = 0.8125$$

**Perhitungan Precision**

$$Precision\ A = \frac{TP}{(TP+FP)} = \frac{18}{(18+10)} = 0.6428$$

$$Precision\ B = \frac{TP}{(TP+FP)} = \frac{14}{(14+3)} = 0.8235$$

$$Precision\ C = \frac{TP}{(TP+FP)} = \frac{13}{(13+1)} = 0.9285$$

$$Precision\ D = \frac{TP}{(TP+FP)} = \frac{20}{(20+1)} = 0.9523$$

$$Precision\ Semua\ Kelas = \frac{0.6428 + 0.8235 + 0.9285 + 0.9523}{4} = 0.8368$$

**Perhitungan Recall**

$$Recall\ A = \frac{TP}{(TP+FN)} = \frac{18}{(18+2)} = 0.9$$

$$Recall\ B = \frac{TP}{(TP+FN)} = \frac{14}{(14+6)} = 0.7$$

$$Recall\ C = \frac{TP}{(TP+FN)} = \frac{13}{(13+7)} = 0.65$$

$$Recall\ D = \frac{TP}{(TP+FN)} = \frac{20}{(20+0)} = 1$$

$$Recall\ Semua\ Kelas = \frac{0.9+0.7+0.65+1}{4} = 0.8125$$



**Perhitungan F1 - Score**

$$F1 A = \frac{2x Precision x Recall}{Precision + Recall} = \frac{2 x 0.6428 x 0.9}{0.6428 + 0.9} = 0.7499$$

$$F1 B = \frac{2x Precision x Recall}{Precision + Recall} = \frac{2 x 0.8235 x 0.7}{0.8235 + 0.7} = 0.7567$$

$$F1 C = \frac{2x Precision x Recall}{Precision + Recall} = \frac{2 x 0.9285 x 0.65}{0.9285 + 0.65} = 0.7646$$

$$F1 D = \frac{2x Precision x Recall}{Precision + Recall} = \frac{2 x 0.9523 x 1}{0.9523 + 1} = 0.9755$$

$$F1 Semua Kelas = \frac{0.7499 + 0.7567 + 0.7646 + 0.9755}{4} = 0.8117$$

Hasil perhitungan secara manual dimasukkan ke dalam Tabel 2 untuk memudahkan pembacaan

**Tabel 2.** Hasil Evaluasi

	Nilai
Accuracy	81.25%
Precision	83.68%
Recall	81.25%
F1-Score	81.17%

Mengacu pada Tabel 2, dapat dilihat masing-masing hasil evaluasi dari klasifikasi genre anime menggunakan KNN. Dari hasil tersebut didapatkan akurasi sebesar 81.25%, Precision sebesar 83.68%, Recall sebesar 81.25%, dan F1-Score sebesar 81.17%.

**5. KESIMPULAN DAN SARAN**

Hasil evaluasi menunjukkan bahwa penelitian ini telah menghasilkan hasil yang cukup baik, dengan akurasi sebesar 81.25%, precision sebesar 83.68%, recall sebesar 81.25%, dan F1-Score sebesar 81.17%. Kesalahan atau kegagalan dalam klasifikasi terjadi karena kemiripan antara kata-kata yang membentuk suatu kelas dengan kata-kata dalam kelas lain serta kata-kata dari setiap data yang digunakan. Evaluasi ini menunjukkan bahwa klasifikasi KNN mampu mengklasifikasikan genre anime dengan baik berdasarkan sinopsis yang ada dalam dataset. Hasil evaluasi menunjukkan kinerja yang memadai, Hasil penelitian ini dapat membentuk landasan yang kuat untuk mengembangkan analisis genre lebih lanjut dengan menggunakan teks sinopsis berbahasa Inggris.

Untuk penelitian berikutnya, dapat dieksplorasi penggunaan algoritma selain KNN, seperti Support Vector Machine (SVM) atau algoritma Naive Bayes, guna menentukan algoritma yang memberikan hasil terbaik dalam klasifikasi genre anime berdasarkan sinopsis, selain itu pula dapat melakukan seleksi fitur agar hanya menggunakan fitur-fitur terbaik dalam melakukan klasifikasi genre berdasarkan sinopsis. Hal ini bertujuan untuk melakukan eksperimen yang dapat mencapai akurasi dan hasil yang lebih baik.

**DAFTAR PUSTAKA:**

[1] M. Aziz and S. Ong, "The Implementation of Japanese Animation (Anime) In Advertising," *J. Indones. Sos. Sains*, vol. 4, no. 04, pp. 370-383, 2023, doi: 10.59141/jiss.v4i05.810.

[2] I. Bondebjerg, "Film: Genres and Genre Theory," *Int. Encycl. Soc. Behav. Sci. Second Ed.*, no. December, pp. 160-164, 2015, doi: 10.1016/B978-0-08-097086-8.95052-9.

[3] P. Putra, A. M. H. Pardede, and S. Syahputra, "Analisis Metode K-Nearest Neighbour (Knn) Dalam Klasifikasi Data Iris Bunga," *J. Tek. Inform. Kaputama*, vol. 6, no. 1, pp. 297-305, 2022.

[4] A. P. Permana, K. Ainiyah, and K. F. H. Holle, "Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol.



- 6, no. 3, pp. 178–188, 2021, doi: 10.14421/jiska.2021.6.3.178-188.
- [5] M. Fahry, “Klasifikasi Multi-Genre Film berdasarkan Sinopsis menggunakan Naive Bayesian Multi-label (NBML) Classifier,” Universitas Telkom, S1 Teknik Informatika, Bandung, 2016. Accessed: Nov. 23, 2023. [Online]. Available: <https://repository.telkomuniversity.ac.id/home/catalog/id/116990/slug/klasifikasi-multi-genre-film-berdasarkan-sinopsis-menggunakan-naive-bayesian-multi-label-nbml-classifier.html>
- [6] A. C. Saputra, A. B. Sitepu, Stanley, P. W. P. Yohanes Sigit, P. G. Sarto Aji Tetuko, and G. C. Nugroho, “The Classification of the Movie Genre based on Synopsis of the Indonesian Film,” *Proceeding - 2019 Int. Conf. Artif. Intell. Inf. Technol. ICAIIT 2019*, pp. 201–204, 2019, doi: 10.1109/ICAIIIT.2019.8834606.
- [7] B. Falakhi, I. Cholissodin, and R. S. Perdana, “Klasifikasi Sinopsis Novel berdasarkan Jenis Genre menggunakan Multi-class Support Vector Machine dan Chi-square,” vol. 7, no. 1, pp. 192–202, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] G. Portolese and V. D. Feltrin, “On the Use of Synopsis-based Features for Film Genre Classification,” pp. 892–902, 2019, doi: 10.5753/eniac.2018.4476.
- [9] D. K. Salamoon, “Anime Sebagai Media Edukasi Digital Mengenai Fungsi Sel Darah Merah (Analisis Visualisasi Karakter AE 3803 Pada Anime Hataraku Saibou),” *Mudra J. Seni Budaya*, vol. 36, no. 2, pp. 197–203, 2021, doi: 10.31091/mudra.v36i2.1468.
- [10] N. Andari and A. G. M. Litaay, “Anime Trend as An Advert Genre in Indonesian Youtube,” *J. Dev. Res.*, vol. 6, no. 2, pp. 164–170, 2022, doi: 10.28926/jdr.v6i2.245.
- [11] M. AbuKausar, V. S. Dhaka, and S. Kumar Singh, “Web Crawler: A Review,” *Int. J. Comput. Appl.*, vol. 63, no. 2, pp. 31–36, 2013, doi: 10.5120/10440-5125.
- [12] A. Y. Muniar, P. Pasnur, and K. R. Lestari, “Penerapan Algoritma K-Nearest Neighbor pada Pengklasifikasian Dokumen Berita Online,” *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 2, p. 137, 2020, doi: 10.35585/inspir.v10i2.2570.
- [13] Dwi Hartanti, Kusriani, and E. L. Taufiq, “PENERAPAN NAÏVE BAYES DALAM PREDIKSI KETERCAPAIAN NILAI KRITERIA KETUNTASAN MINIMAL SISWA JUSIKOM PRIMA ( Junal Sistem Informasi Ilmu Komputer Prima ),” *Jusikom Prima*, vol. 2, no. 1, pp. 15–22, 2018.
- [14] S. Herlambang, S. Basuki, D. R. Akbi, and Z. Sari, “Deteksi Malware Android Berdasarkan System Call Menggunakan Algoritma Support Vector Machine,” vol. 5, pp. 157–165, 2015.
- [15] M. T. Nitamia and H. Februariyanti, “ANALISIS SENTIMEN ULASAN EKSPEDISI J&T EXPRESS MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *MISI(Jurnal Manaj. Inform. Sist. Informasi)*, vol. 5, 2022, doi: <https://doi.org/10.36595/misi.v5i1.396>.
- [16] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*. Springer Cham, 2014. doi: <https://doi.org/10.1007/978-3-319-10247-4>.
- [17] I. W. D. Pancane and I. W. Suriana, “Klasifikasi Komentar Publik Terhadap Kebijakan Pemerintah Pada Facebook Frontpage KOMPAS Menggunakan Clustering K-Means, Furthest First,” *MISI(Jurnal Manaj. Inform. Sist. Informasi)*, vol. 3, no. 2, pp. 166–173, 2020, doi: <https://doi.org/10.36595/misi.v3i2.185>.
- [18] A. F. Hidayatullah, “Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 1, pp. 1–4, 2016, doi: 10.14421/jiska.2016.11-01.
- [19] D. S. Maylawati, W. B. Zulfikar, C. Slamet, M. A. Ramdhani, and Y. A. Gerhana, “An Improved of Stemming Algorithm for Mining Indonesian Text with Slang on Social Media,” *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, 2019, doi: 10.1109/CITSM.2018.8674054.
- [20] S. W. Kim and J. M. Gil, “Research paper classification systems based on TF-IDF and LDA schemes,” *Human-centric Comput. Inf. Sci.*, vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0192-7.
- [21] R. A. Sasmita and A. Z. Falani, “Pemanfaatan Algoritma TF/IDF pada



sistem Informasi Ecomplaint Handling,” *J. Ilm. Lintas Sist. Inf. dan Komput.*, vol. 27, no. 1, pp. 27–33, 2018, [Online]. Available: <https://jurnal.narotama.ac.id/index.php/link/article/view/1700>

[22] A. K. Nikhath, K. Subrahmanyam, and R. Vasavi, “Building a K-Nearest Neighbor Classifier for Text Categorization,” *Int. J. Comput. Sci. Inf. Technol.*, 2016.