

ANALISIS POPULARITAS GENRE FILM DI INDONESIA MENGUNAKAN ALGORITMA DECISION TREE

**Freyro Dobry Sianipar, Ananda Irya Shakila Syukron, Aqilah Defiyanti,
Alvin Hafiz, Revidamurti Dly, Fanny Ramadhani, Arnita**

Prodi Ilmu Komputer, Fakultas MIPA, Universitas Negeri Medan Jl. William Iskandar Ps. V, Medan Estate,
Kec. Percut Sei Tuan Kabupaten Deli Serdang, Sumatera Utara 20222, Sumatera Utara
freyrodobry31@gmail.com

ABSTRAK

Industri perfilman di Indonesia mengalami pertumbuhan pesat dengan semakin beragamnya genre yang diminati oleh penonton. Namun, pola popularitas genre film masih menjadi tantangan dalam pengambilan keputusan industri. Penelitian ini bertujuan untuk menganalisis faktor yang memengaruhi popularitas genre film di Indonesia dengan menerapkan algoritma Decision Tree. Data yang digunakan berasal dari The Movie Database (TMDb) mencakup 3.000 film dari tahun 2015 hingga 2024, dengan variabel utama berupa jumlah penonton dan rating. Proses analisis mencakup tahapan preprocessing data, transformasi data, serta penyeimbangan kelas menggunakan teknik Synthetic Minority Over-sampling Technique (SMOTE). Hasil penelitian menunjukkan bahwa rating merupakan faktor dominan dalam menentukan popularitas genre film, diikuti oleh genre Romance, Drama, dan Horror. Model Decision Tree yang diterapkan memiliki akurasi sebesar 72,98% dalam mengklasifikasikan film populer dan tidak populer. Temuan ini memberikan kontribusi terhadap pemahaman preferensi penonton di Indonesia dan dapat menjadi acuan bagi industri perfilman dalam menyusun strategi produksi dan pemasaran yang lebih efektif. Berbeda dengan penelitian sebelumnya yang berfokus pada sistem rekomendasi atau analisis faktor individu, penelitian ini menawarkan pendekatan klasifikasi genre film berdasarkan kombinasi jumlah penonton dan rating, sehingga memberikan perspektif yang lebih komprehensif dalam mengidentifikasi tren popularitas film di Indonesia.

Kata kunci : Popularitas Film, Genre Film, Data Mining, Decision Tree, TMDb, SMOTE.

1. PENDAHULUAN

Industri perfilman di Indonesia mengalami pertumbuhan yang sangat signifikan, ditandai dengan semakin beragamnya genre film yang dirilis setiap tahunnya. Seiring dengan hal tersebut, preferensi penonton terhadap genre tertentu juga terus berkembang. Kesuksesan suatu film kini tidak hanya bergantung pada jumlah penonton di bioskop, tetapi juga dipengaruhi oleh faktor lain seperti rating dan popularitas film [1].

Meskipun demikian, ketidakpastian mengenai keberhasilan sebuah film menjadikan industri ini memiliki tingkat risiko yang tinggi. Oleh karena itu, kemampuan untuk memprediksi kesuksesan film sangat bermanfaat bagi berbagai pihak yang terlibat, seperti rumah produksi dan investor, dalam upaya meminimalkan potensi kerugian. Jika komputer dapat digunakan untuk memperkirakan tingkat keberhasilan suatu film sebelum dirilis, hal ini akan memberikan manfaat besar bagi industri perfilman [2].

Dengan melimpahnya data terkait film, penerapan data Mining menjadi solusi yang tepat untuk mengidentifikasi pola dan tren dalam preferensi penonton. Salah satu metode yang banyak digunakan untuk menganalisis dan mengklasifikasikan data adalah algoritma *Decision Tree*, yang mampu menentukan faktor dominan yang mempengaruhi tingkat popularitas suatu genre film. Algoritma ini sangat cocok untuk menentukan apakah suatu film termasuk dalam kategori populer atau tidak, berdasarkan jumlah penonton dan rating yang

diterima. Selain itu, *Decision Tree* memiliki keunggulan dalam hal interpretasi yang mudah, sehingga memungkinkan visualisasi keputusan dalam bentuk pohon yang jelas dan mudah dipahami oleh para pelaku industri film. Dengan pendekatan ini, penelitian dapat mengidentifikasi faktor-faktor utama yang mempengaruhi popularitas film, seperti rating dan genre tertentu, termasuk Romance, Drama, dan Horror. Data diperoleh dari The Movie Database (TMDb), mencakup 3.000 film yang dirilis dalam rentang waktu 2015 hingga 2024. Analisis akan difokuskan pada jumlah penonton dan rating, sebagai indikator utama dalam mengukur popularitas genre film.

Beberapa penelitian sebelumnya telah meneliti topik serupa, seperti penelitian yang dilakukan oleh [3], membahas tentang implementasi teknik data mining, khususnya algoritma Apriori, untuk mengidentifikasi pola dalam elemen film yang disukai oleh penonton. Penelitian ini bertujuan untuk memahami preferensi penonton terhadap elemen film seperti alur cerita, genre, dan karakter, yang dapat membantu produser dalam menciptakan film yang menarik. Hasilnya menunjukkan bahwa algoritma Apriori berhasil menemukan delapan aturan asosiasi yang menghubungkan elemen film seperti alur cerita, genre, dan karakter dengan preferensi penonton. Selain itu, penelitian oleh [4], menerapkan algoritma *Decision Tree* dalam sistem rekomendasi film, yang membantu meningkatkan pengalaman pengguna di platform Netflix. Hasil dari penelitian ini

menunjukkan bahwa algoritma Decision Tree berhasil meningkatkan sistem rekomendasi dan klasifikasi rating di platform Netflix.

Berbeda dengan penelitian sebelumnya, penelitian ini menerapkan algoritma Decision Tree untuk mengklasifikasikan popularitas genre film di Indonesia dengan mempertimbangkan kombinasi jumlah penonton dan rating sebagai faktor utama. Penelitian ini bertujuan untuk mengatasi ketidakpastian yang dihadapi industri perfilman terkait genre film yang paling diminati oleh penonton. Melalui analisis data dari The Movie Database (TMDb), penelitian ini berusaha mengidentifikasi faktor-faktor kunci yang mempengaruhi popularitas film, sehingga dapat memberikan wawasan berharga bagi rumah produksi, investor, dan platform streaming dalam pengambilan keputusan strategis. Selain itu, penelitian ini juga bertujuan untuk mengembangkan model prediksi yang dapat memetakan tren popularitas film berdasarkan data historis, memungkinkan industri film merancang strategi produksi dan pemasaran yang lebih efektif.

2. TINJAUAN PUSTAKA

Dalam penelitian sebelumnya [4], Menjelaskan penerapan algoritma *Decision Tree* dalam sistem rekomendasi film di platform streaming. Penelitian ini menemukan bahwa Decision Tree mampu meningkatkan akurasi rekomendasi dengan mengelompokkan film berdasarkan preferensi pengguna. Dengan mempertimbangkan variabel seperti rating, genre, serta histori tontonan, model ini memberikan prediksi yang lebih personal terhadap film yang sesuai untuk setiap pengguna. Dari penelitian sebelumnya, menggunakan pendekatan yang masih berfokus pada sistem rekomendasi atau analisis faktor individu.

Berbeda dengan penelitian sebelumnya, penelitian ini menawarkan pendekatan dengan mengimplementasikan algoritma Decision Tree untuk mengklasifikasikan popularitas genre film di Indonesia berdasarkan jumlah penonton dan rating.

2.1. Data Mining dalam Analisis Film

Data mining merupakan proses ekstraksi informasi yang berharga dari kumpulan data besar dan kompleks dengan tujuan menghasilkan informasi baru yang dapat digunakan dalam pengambilan keputusan.

Dalam konteks industri film, data mining digunakan untuk menganalisis faktor-faktor yang memengaruhi popularitas suatu film. Proses ini mencakup pengumpulan, pengolahan, dan pemodelan data guna mengidentifikasi pola atau tren yang dapat menjelaskan maupun memprediksi kesuksesan sebuah film. Data yang dianalisis mencakup berbagai aspek, seperti rating penonton, jumlah penonton, ulasan, genre, anggaran produksi, serta faktor eksternal lainnya, seperti strategi pemasaran dan tren sosial.

Berbagai teknik data mining, seperti klasifikasi, klustering, dan analisis asosiasi, dapat diterapkan

dalam analisis popularitas film. Teknik klasifikasi, misalnya, dapat digunakan untuk memprediksi rating film berdasarkan genre dan elemen lainnya. Sementara itu, klustering dapat membantu mengidentifikasi kelompok audiens dengan preferensi yang serupa, sehingga memungkinkan industri film untuk merancang strategi pemasaran yang lebih tepat sasaran [5].

2.2. Popularitas Genre Film

Popularitas film dapat diukur berdasarkan berbagai faktor, seperti rating, jumlah penonton, serta ulasan dari kritikus dan penonton umum. Film yang populer tentu dapat dikaitkan dengan kombinasi antara alur cerita yang menarik, kualitas produksi yang baik, serta strategi pemasaran yang efektif [6]. Preferensi penonton terhadap genre film terus mengalami perubahan, terutama di era media sosial. Beberapa genre tetap populer karena memiliki daya Tarik universal, sementara genre lain mendapatkan perhatian lebih besar seiring dengan tren yang berkembang [7].

Dengan semakin beragamnya genre film di Indonesia, pengalaman menonton bagi setiap orang pun bisa berbeda. Setiap genre memiliki cara tersendiri dalam menyampaikan cerita, yang bisa memengaruhi bagaimana Visual Reader menjelaskan adegan kepada penonton. Hal ini dapat membuat pemahaman penonton menjadi lebih mudah atau justru lebih menantang, tergantung pada kompleksitas visual yang ada dalam film tersebut [8].

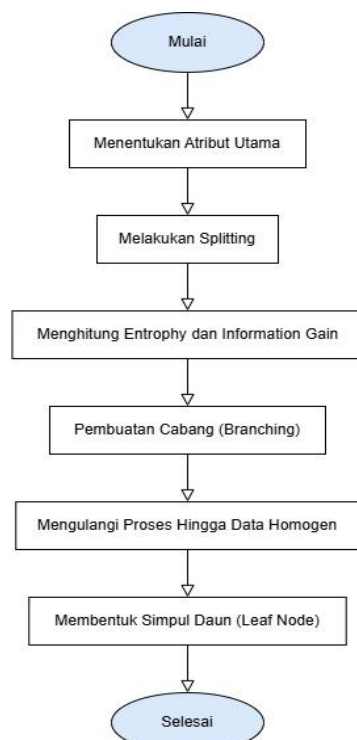
2.3. The Movie Database (TMDb)

TMDb (The Movie Database) adalah platform yang menyediakan informasi film secara real-time, termasuk sinopsis, rating, dan genre. Setiap individu tentunya memiliki preferensi yang berbeda dalam mencari hiburan mereka, seperti menonton film, untuk mengalihkan perhatian dari kesibukan, menghilangkan rasa bosan bahkan hanya sebagai rutinitas harian. Untuk itu agar dapat menyesuaikan beragam preferensi pengguna, diperlukan sistem informasi yang dapat merekomendasikan film sesuai kebutuhan, minat dan selera pengguna baik untuk hiburan maupun tujuan profesional [9].

2.4. Klasifikasi

Klasifikasi adalah proses mengelompokkan data berdasarkan karakteristik tertentu agar lebih terstruktur dan mudah digunakan. Dalam industri film, klasifikasi genre berfungsi untuk mengelompokkan film berdasarkan tema, alur cerita, maupun elemen khas lainnya, sehingga membantu penonton menemukan film sesuai preferensi mereka. Namun, disebabkan oleh banyaknya film yang mengandung lebih dari satu genre, proses klasifikasi sering kali menjadi kompleks dan memerlukan metode yang akurat untuk memastikan setiap film masuk dalam kategori yang paling sesuai [10].

2.5. Algoritma Decision Tree



Gambar 1. Alur Decision Tree

Decision Tree adalah metode pemodelan prediktif dalam analisis data yang menggunakan struktur berbentuk pohon untuk menggambarkan proses pengambilan keputusan berdasarkan serangkaian aturan dan kondisi. Metode ini sering digunakan dalam machine learning untuk melakukan klasifikasi dan regresi dengan mempartisi dataset secara rekursif [4]. [11] mendefinisikan decision tree sebagai proses klasifikasi suatu sampel data yang belum diketahui kelasnya ke dalam kelas-kelas yang telah ada.

Untuk meningkatkan akurasi, metode ini memanfaatkan teknik bagging dan boosting karena rentan terhadap overfitting. Algoritma ini memiliki struktur menyerupai pohon [12].

Decision Tree bekerja seperti proses pengambilan keputusan bertahap. Prosesnya dimulai dari akar, lalu berlanjut dengan mengevaluasi suatu fitur dan memilih salah satu dari dua cabang yang tersedia. Proses ini terus berlanjut hingga mencapai bagian akhir, yaitu daun, yang biasanya menunjukkan hasil akhir atau target yang dicari. Salah satu keunggulan Decision Tree adalah ketahanannya terhadap variasi nilai dalam fitur data, sehingga tetap bisa bekerja dengan baik meskipun datanya belum dinormalisasi. Dalam Scikit-learn, algoritma ini dapat dilatih menggunakan ukuran Gini atau cross-entropy impurity untuk menentukan tingkat keakuratannya [13].

Decision Tree terdiri dari simpul akar (root node) yang merepresentasikan fitur utama, simpul keputusan (internal nodes) yang menunjukkan pengujian suatu

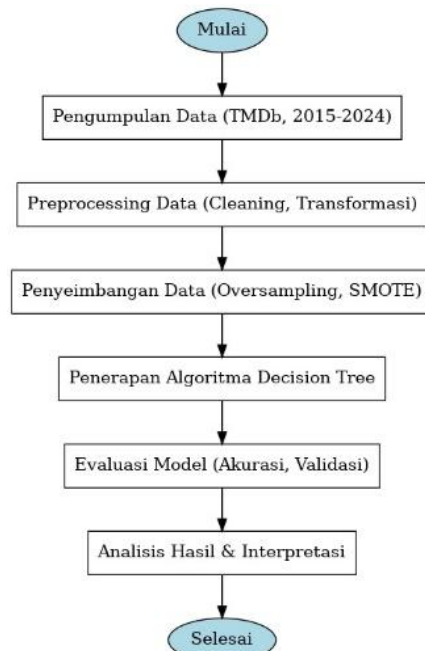
atribut, cabang (branches) sebagai hasil dari pengujian, dan simpul daun (leaf nodes) yang menunjukkan kategori akhir atau hasil keputusan [14].

Dalam analisis popularitas genre film, Decision Tree memungkinkan peneliti untuk mengidentifikasi faktor-faktor dominan yang memengaruhi popularitas suatu genre di Indonesia. Dengan memetakan hubungan antara jumlah penonton, rating, dan faktor lainnya, model ini dapat membantu memprediksi genre yang memiliki potensi popularitas tinggi di masa depan.

Decision Tree sering diterapkan dalam berbagai bidang, seperti pengenalan pola, prediksi, dan analisis keputusan karena modelnya mudah dipahami dan diinterpretasikan. Dalam proses klasifikasi, algoritma *Decision Tree* bekerja dengan memecah dataset menjadi kelompok-kelompok yang lebih kecil berdasarkan atribut yang ada, hingga setiap kelompok hanya berisi satu kelas yang sama. Proses ini dikenal sebagai *splitting* atau pembagian data. Langkah-langkah pembuatan *Decision Tree*:

- Menentukan atribut utama: Atribut yang paling berpengaruh dalam klasifikasi dipilih sebagai *root node*.
- Melakukan *splitting*: Dataset dibagi menjadi beberapa subset berdasarkan atribut utama yang telah dipilih.
- Menghitung *Entropy* dan *Information Gain*: *Entropy* digunakan untuk mengukur tingkat ketidakpastian dalam dataset. Semakin rendah entropi, semakin baik pembagian data. Sedangkan *Information Gain* menghitung seberapa besar pengurangan *entropy* setelah dilakukan pembagian data, sehingga membantu dalam memilih atribut terbaik untuk *splitting* selanjutnya [15].

3. METODE PENELITIAN



Gambar 2. Alur Penelitian

Penelitian ini dilakukan secara kuantitatif dengan metode Data Mining untuk menganalisis popularitas genre film berdasarkan rating, dan jumlah penonton. Penelitian ini bersifat deskriptif dan eksploratif, di mana data yang telah dikumpulkan akan dianalisis menggunakan teknik klasifikasi untuk menemukan pola dari genre film yang paling diminati oleh penonton Indonesia.

Metode Decision Tree digunakan untuk mengidentifikasi faktor utama yang mempengaruhi popularitas suatu genre berdasarkan rating dan jumlah penonton. Adapun tahapan penelitian ini adalah sebagai berikut:

3.1. Data dan sumber data

Penelitian ini menggunakan data yang berasal dari The Movie Database (TMDb). Data yang digunakan adalah data film dengan rentang tahun 2015 hingga 2024. Data diperoleh dengan mengakses API TMDb (web scrapping), sehingga diperoleh data berjumlah 3000 film dan 5 atribut (fitur). Data ini akan diolah menggunakan Bahasa pemrograman Python dengan library pandas untuk manipulasi data, scikit-learn untuk pemodelan machine learning, dan imbalanced-learn untuk menangani data yang tidak seimbang.

Tabel 1. Dataset Penelitian

Jenis Data	Jmlah Data
Film	3000
Atribut	10

Tabel 2. Jenis Data Atribut

Jenis Data	Jmlah Data
1	Judul
2	Tahun
3	Genre
4	Rating
5	Vote_Count
6	Pendapatan
7	Anggaran
8	Durasi
9	Popularitas
10	Ulasan

3.2. Preprocessing Data

Pada tahap peneltian ini dilakukan 3 tahap yaitu pembersihan data, transformasi data, dan penanganan ketidakseimbangan data.

a. Pembersihan Data

Di tahap pembersihan data, beberapa langkah dilakukan untuk menghapus data duplikat untuk menghindari pengaruh berlebih dari satu film yang tercatat lebih dari sekali dan juga memeriksa dan menangani nilai yang hilang atau tidak valid dalam dataset tersebut. Selanjutnya, genre dengan jumlah sampel yang sangat sedikit (kurang dari 10 film) dihapus untuk menghindari ketidakseimbangan data yang ekstrem.

b. Transformasi Data

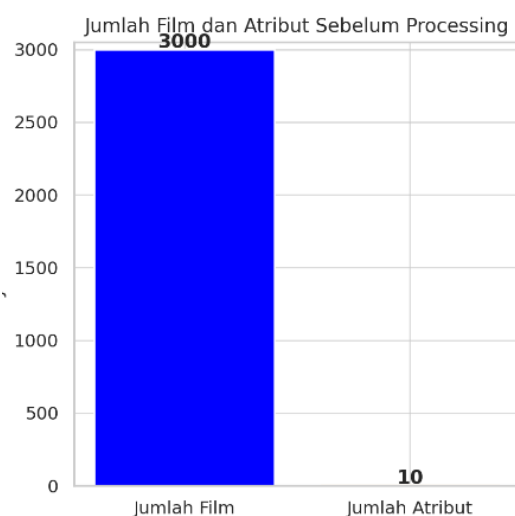
Pada tahap transformasi data, genre film dalam dataset diubah menjadi fitur karena satu film

sering kali memiliki lebih dari satu genre. Dalam penelitian ini, genre digunakan sebagai fitur biner, di mana setiap genre akan diberi nilai 1 jika ada pada film tersebut dan 0 jika tidak ada. Pendekatan ini dipilih karena memungkinkan kita untuk menganalisis bagaimana faktor-faktor seperti rating dan popularitas memengaruhi popularitas berbagai genre film. Dengan cara ini, model machine learning dapat lebih mudah mengidentifikasi hubungan antara genre dan atribut lainnya, seperti rating atau jumlah penonton, untuk melihat faktor-faktor apa yang paling berpengaruh dalam menentukan popularitas genre. Dengan menjadikan genre sebagai fitur, bukan target, penelitian ini juga dapat mengakomodasi fakta bahwa setiap film dapat memiliki lebih dari satu genre, yang akan menghasilkan pemodelan yang lebih tepat dan representatif.

c. Penyeimbangan Data

Pada tahap penelitian ini, akan melakukan penanganan ketidakseimbangan kelas terhadap kelas minoritas menggunakan teknik resampling yaitu oversampling. Teknik oversampling dipilih karena tidak mengurangi dataset akan tetapi menambah dataset yang kurang pada kelas minoritas [16].

Oversampling yang digunakan adalah Synthetic Minority Over-sampling Technique (SMOTE). SMOTE digunakan untuk menambah sampel pada kelas yang memiliki jumlah data lebih sedikit dengan cara mensintesis data baru berdasarkan sampel yang ada. Proses ini dilakukan dengan mencari titik-titik tetangga terdekat dari sampel minoritas, kemudian menghasilkan data baru di antara titik-titik tersebut. Dengan demikian, distribusi kelas dalam dataset menjadi lebih seimbang, sehingga model machine learning yang diterapkan tidak cenderung hanya mengenali kelas mayoritas.



Gambar 3. Dataset sebelum processing

3.3. Penerapan Algoritma Decision Tree

Penerapan algoritma Decision Tree dalam penelitian ini bertujuan untuk mengklasifikasikan pola popularitas genre film berdasarkan data yang tersedia. Algoritma ini bekerja dengan membangun pohon keputusan berdasarkan fitur-fitur yang mempengaruhi popularitas film, seperti rating dan jumlah penonton. Proses ini dimulai dengan pemilihan atribut yang paling berpengaruh sebagai akar pohon, diikuti dengan pemisahan data ke dalam cabang-cabang berdasarkan nilai atribut tersebut. Dengan pendekatan ini, pola klasifikasi dapat diperoleh secara sistematis dan dapat digunakan untuk memahami faktor-faktor yang mendukung keberhasilan suatu genre film di Indonesia.

3.4. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja algoritma Decision Tree dalam melakukan klasifikasi popularitas genre film. Pengukuran ini dilakukan dengan metode validasi, seperti confusion matrix, akurasi, presisi, recall, dan F1-score, guna memastikan model yang digunakan memiliki tingkat keandalan yang baik. Selain itu, dilakukan pengujian menggunakan data uji yang terpisah dari data latih untuk mengetahui sejauh mana model dapat melakukan prediksi terhadap data yang belum dikenali. Hasil evaluasi ini menjadi dasar dalam menentukan efektivitas model yang diterapkan dalam penelitian.

3.5. Analisis Hasil dan Interpretasi

Setelah model dievaluasi, langkah selanjutnya adalah menganalisis hasil yang diperoleh serta memberikan interpretasi terhadap pola yang ditemukan. Analisis ini dilakukan dengan membandingkan hasil prediksi dengan data aktual guna mengetahui hubungan antara faktor-faktor tertentu dengan popularitas genre film. Selain itu, pola yang dihasilkan oleh Decision Tree dapat digunakan untuk memahami tren film yang diminati oleh penonton di Indonesia. Dengan demikian, penelitian ini dapat memberikan wawasan yang berguna bagi industri perfilman dalam mengembangkan strategi produksi yang sesuai dengan preferensi pasar.

4. HASIL DAN PEMBAHASAN

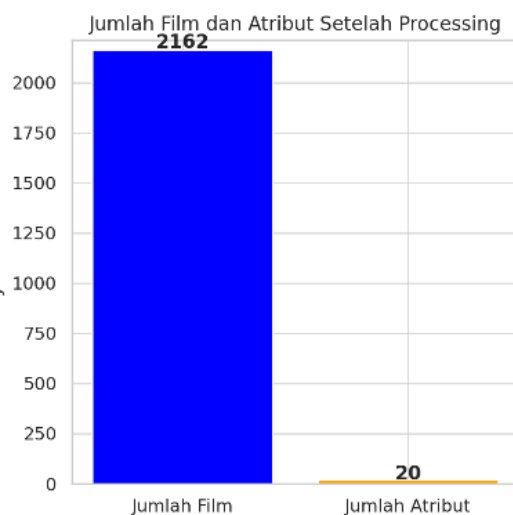
4.1. Hasil Preprocessing

Setelah dilakukan tahap preprocessing, dataset yang digunakan dalam penelitian ini mengalami beberapa perubahan signifikan. Dari total 3000 data dan 10 atribut film yang dikumpulkan dari The Movie Database (TMDb) dengan rentang tahun 2015 hingga 2024, dilakukan pembersihan data dengan menghapus duplikasi dan menangani nilai yang hilang atau tidak valid. Selain itu, genre dengan jumlah sampel yang sangat sedikit (kurang dari 10 film) juga dihapus untuk menghindari ketidakseimbangan yang ekstrem dalam distribusi kelas. Setelah tahap ini, jumlah total data film yang digunakan dalam analisis berkurang menjadi

2162 film dengan 20 atribut. Proses transformasi data juga dilakukan dengan mengubah genre film menjadi fitur biner, di mana setiap film dapat memiliki lebih dari satu genre secara simultan. Dengan demikian, dataset hasil preprocessing telah lebih siap untuk dianalisis dalam model machine learning, karena telah mengurangi kemungkinan bias akibat data duplikat dan telah mempertimbangkan distribusi genre yang lebih representatif.

Tabel 3. Jenis Data Atribut

Jenis Data	Jmlah Data
1	Judul
2	Tahun
3	Popularitas
4	Rating
5	Action
6	Adventure
7	Animation
8	Crime
9	Dokumenter
10	Drama
11	Family
12	Fantasy
13	Fiction
14	History
15	Horror
16	Comedy
17	Music
18	Mystery
19	Romance
20	Thriller



Gambar 4. Dataset setelah processing

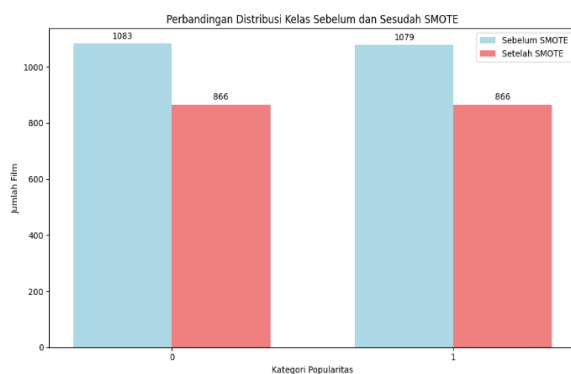
4.2. Hasil Penyeimbangan Data

Pada tahap ini, dilakukan proses penyeimbangan data menggunakan metode *Synthetic Minority Over-sampling Technique (SMOTE)* guna mengatasi ketidakseimbangan jumlah sampel dalam dataset. Sebelum penerapan SMOTE, jumlah data dalam kategori tidak populer (0) sebanyak 1.083 film, sedangkan jumlah data dalam kategori populer (1) sebanyak 1.079 film. Meskipun perbedaan ini tidak

terlalu signifikan, ketidakseimbangan ini tetap berpotensi memengaruhi kinerja model dalam melakukan klasifikasi. Oleh karena itu, SMOTE diterapkan untuk menambah sampel pada kelas minoritas dengan cara mensintesis data baru berdasarkan titik-titik yang telah ada. Setelah proses ini, jumlah data dalam kedua kategori menjadi lebih seimbang, yakni 866 film untuk masing-masing kelas. Gambar 3 yang disajikan menunjukkan perbandingan distribusi sebelum dan sesudah penerapan SMOTE, di mana distribusi kelas menjadi lebih merata.

Dengan adanya distribusi data yang lebih seimbang, model klasifikasi yang diterapkan dalam penelitian ini diharapkan dapat bekerja lebih optimal tanpa adanya kecenderungan untuk lebih mengutamakan kelas mayoritas. Namun, penting untuk diperhatikan bahwa SMOTE tidak menambah informasi baru dalam dataset, melainkan hanya menciptakan variasi dari data yang sudah ada. Oleh karena itu, meskipun metode ini efektif dalam mengatasi ketidakseimbangan data, terdapat kemungkinan bahwa model akan mengalami overfitting jika terlalu bergantung pada sampel sintesis yang dihasilkan.

Untuk mengurangi risiko overfitting, dilakukan validasi model menggunakan teknik cross-validation serta pengujian terhadap data uji yang tidak termasuk dalam proses pelatihan. Selain itu, pemilihan fitur yang tepat serta tuning parameter model juga menjadi langkah penting dalam memastikan bahwa model yang digunakan dapat melakukan generalisasi dengan baik terhadap data baru. Dengan demikian, penerapan SMOTE dalam penelitian ini telah berhasil menyeimbangkan distribusi kelas dalam dataset.



Gambar 5. Perbandingan sebelum dan sesudah smote

4.3. Hasil Pengujian

Pengujian dilakukan terhadap model klasifikasi yang dikembangkan menggunakan algoritma Decision Tree. Dataset yang digunakan berasal dari TMDb dan telah melalui tahap preprocessing untuk memastikan kualitas data. Model ini dibangun dengan mempertimbangkan beberapa atribut utama seperti rating, jumlah penonton, dan faktor lain yang memengaruhi popularitas genre film. Dalam proses pelatihan, dataset dibagi menjadi data latih dan data uji agar performa model dapat diukur secara akurat.

Algoritma Decision Tree dalam pengujian ini menggunakan beberapa perhitungan matematis untuk menentukan pemisahan data yang optimal. Salah satu perhitungan utama adalah entropy, yang digunakan untuk mengukur tingkat ketidakpastian dalam dataset. Entropy dihitung dengan rumus:

$$H(S) = -\sum \pi \log_2 \pi \quad (1)$$

Information Gain dihitung dengan rumus:

$$IG(S, A) = H(S) - \sum \frac{S_v}{S} H(S_v) \quad (2)$$

Gini Index dihitung dengan rumus:

$$Gini(S) = 1 - \sum \pi^2 \quad (3)$$

Model diuji dengan confusion matrix untuk mendapatkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Rumus akurasi yang digunakan:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

Model Decision Tree yang digunakan dalam penelitian ini menunjukkan tingkat akurasi sebesar 72,98%. Untuk mengevaluasi performa model, digunakan metrik precision, recall, dan F1-score, yang ditampilkan dalam table berikut:

Tabel 4. Evaluasi model

Kelas	Precision	Recall	F1-score	Support
Tidak Populer	0.72	0.75	0.73	216
Populer	0.74	0.71	0.72	217
Rata-rata	0.73	0.73	0.73	433

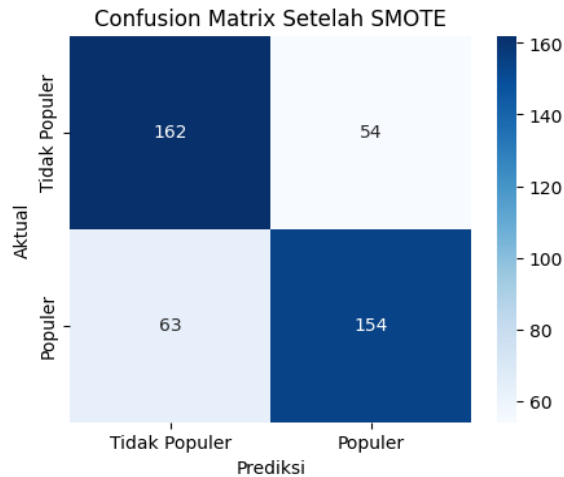
Tabel 5. Akurasi

Akurasi
0.7298

Hasil pengujian menunjukkan bahwa model memiliki precision sebesar 72% untuk kategori Tidak Populer dan 74% untuk kategori Populer, sementara nilai recall sebesar 75% untuk kategori Tidak Populer dan 71% untuk kategori Populer. Nilai F1-score untuk kedua kategori berkisar antara 72% hingga 73%, yang mengindikasikan keseimbangan antara kemampuan model dalam mengidentifikasi film populer maupun tidak populer.

Berdasarkan gambar 5 confusion matrix, model berhasil mengklasifikasikan 162 film tidak populer dengan benar, namun terdapat 54 film tidak populer yang salah diklasifikasikan sebagai populer. Sebaliknya, 154 film populer berhasil diidentifikasi dengan benar, tetapi masih ada 63 film populer yang diklasifikasikan sebagai tidak populer. Hal ini

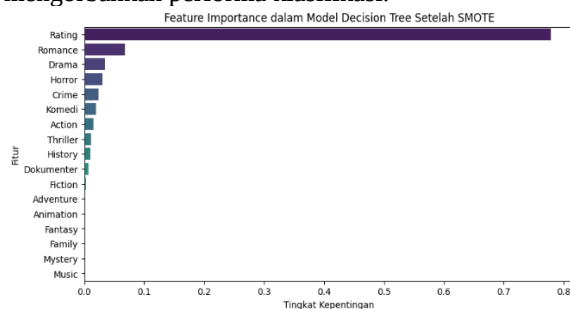
menunjukkan bahwa model cenderung lebih baik dalam mengenali film yang tidak populer dibandingkan dengan mengklasifikasikan film yang benar-benar populer. Untuk meningkatkan performa model, dapat dilakukan validasi silang (cross-validation) guna mengevaluasi kestabilan model di berbagai skenario pembagian data.



Gambar 6. Visualisasi confusion matrix

Dalam gambar 6 analisis feature importance, ditemukan bahwa rating merupakan faktor paling dominan dalam menentukan popularitas film, dengan kontribusi sebesar 77,83%. Hal ini mengindikasikan bahwa semakin tinggi rating yang diberikan oleh penonton, semakin besar kemungkinan film tersebut dikategorikan sebagai populer. Selain rating, genre yang berpengaruh dalam klasifikasi popularitas film adalah Romance (6,83%), Drama (3,46%), Horror (3,00%), dan Crime (2,39%). Keberadaan genre-genre ini menunjukkan bahwa beberapa jenis film memiliki daya tarik lebih besar dalam menarik perhatian penonton.

Namun, beberapa fitur seperti Adventure, Animation, Fantasy, Family, Mystery, dan Music memiliki tingkat kepentingan nol dalam model ini. Artinya, fitur-fitur tersebut tidak memberikan kontribusi dalam menentukan popularitas film dan dapat dihapus dalam pengembangan model selanjutnya. Penghapusan fitur yang tidak berpengaruh dapat meningkatkan efisiensi komputasi serta mengurangi kompleksitas model tanpa mengorbankan performa klasifikasi.



Gambar 7. Visualisasi feature importance

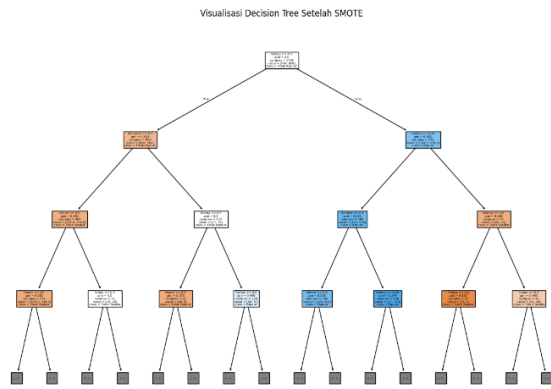
4.4. Pohon Keputusan

Visualisasi pohon keputusan yang ditampilkan pada gambar 7 menunjukkan proses klasifikasi genre film berdasarkan atribut yang mempengaruhi tingkat popularitasnya. Pada simpul akar (root node), atribut rating menjadi faktor utama dalam menentukan apakah suatu film termasuk dalam kategori populer atau tidak populer. Jika nilai rating berada di bawah ambang batas tertentu, model cenderung mengklasifikasikan film sebagai tidak populer. Sebaliknya, jika rating lebih tinggi, kemungkinan besar film tersebut dikategorikan sebagai populer. Pemilihan rating sebagai atribut utama menunjukkan bahwa skor penilaian dari penonton memiliki peran signifikan dalam menentukan daya tarik sebuah film di Indonesia.

Pada percabangan selanjutnya, faktor genre mulai berperan dalam memperjelas klasifikasi. Genre seperti Romance, Drama, dan Komedi memiliki batasan tertentu yang berpengaruh terhadap keputusan model. Sebagai contoh, dalam cabang yang mengarah ke film tidak populer, genre Romance dengan jumlah tertentu menjadi faktor yang memperkuat klasifikasi tersebut. Di sisi lain, dalam cabang yang mengarah ke film populer, genre Komedi dan Drama menunjukkan keterkaitan yang lebih besar dengan film yang memiliki banyak penonton dan rating tinggi. Hal ini mengindikasikan bahwa genre tertentu lebih cenderung menarik minat audiens dibandingkan genre lainnya.

Lebih jauh, model juga mempertimbangkan sub-genre seperti Horror, Action, Crime, dan Thriller dalam proses pengambilan keputusan. Genre Horror yang memiliki nilai di bawah ambang batas tertentu sering kali diklasifikasikan ke dalam kategori film tidak populer. Sebaliknya, genre Thriller yang dikombinasikan dengan rating tinggi lebih sering muncul dalam jalur yang mengarah ke film populer. Keberadaan fitur genre dalam berbagai level pohon menunjukkan bahwa meskipun rating menjadi faktor dominan, preferensi terhadap jenis film tertentu juga memberikan kontribusi yang signifikan terhadap popularitasnya.

Struktur pohon yang terbentuk memperlihatkan pola pengambilan keputusan yang jelas, di mana setiap percabangan mewakili aturan yang semakin spesifik dalam menentukan kategori film. Pohon keputusan ini menunjukkan bahwa faktor rating dan genre merupakan aspek utama dalam menentukan apakah suatu film akan mendapatkan perhatian yang besar dari penonton atau tidak. Dengan demikian, hasil analisis ini dapat dijadikan dasar bagi industri perfilman dalam memahami tren popularitas film serta dalam merancang strategi pemasaran yang lebih efektif berdasarkan karakteristik film yang diminati oleh audiens.



Gambar 8. Visualisasi pohon keputusan

5. KESIMPULAN DAN SARAN

Berdasarkan hasil klasifikasi menggunakan model Decision Tree, tingkat akurasi yang diperoleh adalah 72,98%. Dari confusion matrix yang dihasilkan, model memiliki kecenderungan untuk lebih sering mengklasifikasikan film dengan benar, tetapi masih terdapat kesalahan prediksi terutama dalam membedakan film yang populer dan tidak populer. Hal ini ditunjukkan dengan 54 film yang tidak populer diprediksi sebagai populer, serta 63 film yang sebenarnya populer tetapi diklasifikasikan sebagai tidak populer. Secara keseluruhan, precision dan recall untuk kedua kategori cukup seimbang dengan nilai rata-rata sekitar 73%. Hasil analisis terhadap fitur yang berpengaruh menunjukkan bahwa rating merupakan faktor dominan dalam menentukan popularitas genre film, dengan tingkat kepentingan sebesar 77,83%. Setelah rating, genre yang memiliki pengaruh cukup besar terhadap popularitas adalah Romance (6,83%), Drama (3,46%), Horror (3,00%), dan Crime (2,39%).

Hal ini mengindikasikan bahwa film dengan rating tinggi lebih cenderung diklasifikasikan sebagai populer, sementara genre tertentu seperti Romance dan Drama juga berkontribusi terhadap prediksi popularitas film. Namun, beberapa genre seperti Adventure, Animation, Fantasy, Family, Mystery, dan Music memiliki tingkat kepentingan nol, yang menunjukkan bahwa mereka tidak memiliki pengaruh yang signifikan dalam klasifikasi. Penelitian ini masih memiliki beberapa keterbatasan yang bisa dikembangkan lebih lanjut. Faktor yang dianalisis hanya jumlah penonton dan rating, padahal ada banyak hal lain yang bisa memengaruhi popularitas film, seperti strategi pemasaran atau tren media sosial. Data yang digunakan juga hanya dari TMDb, yang mungkin kurang mewakili selera penonton di Indonesia. Selain itu, akurasi model Decision Tree masih bisa ditingkatkan dengan algoritma lain seperti Random Forest. Metode SMOTE yang digunakan untuk menyeimbangkan data juga belum menambah informasi baru, sehingga teknik lain bisa dicoba. Ke depannya, penelitian bisa lebih fokus pada preferensi individu agar hasilnya lebih relevan bagi industri perfilman.

DAFTAR PUSTAKA

- [1] D. Ariatmanto and M. Ilham Arief, "Prediksi Peluang Kesuksesan Film Dalam Pra Produksi Menggunakan Algoritma Decision Tree," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 222–226, 2023, doi: 10.36040/jati.v7i1.6277.
- [2] H. Athaya, I. T. A. Amrulloh, and R. Mufidah, "Prediksi Kesuksesan Film Indonesia Menggunakan Jaringan Syaraf Tiruan," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 2106–2111, 2024, doi: 10.36040/jati.v8i2.9384.
- [3] I. S. Syahputra, H. -, and A. R. Dewi, "Implementasi Data Mining Dalam Menentukan Pola Elemen Film Favorit Menggunakan Algoritma Apriori," *Syntax J. Softw. Eng. Comput. Sci. Inf. Technol.*, vol. 4, no. 1, pp. 289–294, 2023, doi: 10.46576/syntax.v4i1.2888.
- [4] D. A. Mukhsinin, M. Rafliansyah, S. A. Ibrahim, R. Rahmadden, and D. Wulandari, "Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 570–579, 2024, doi: 10.57152/malcom.v4i2.1255.
- [5] A. Damayanti, "Penerapan Data Mining untuk Analisa Pemilihan Genre Film menggunakan Algoritma Apriori (Studi Kasus : imdb . com) Application of Data Mining for Analysis of Film Genre Selection using a Priori Algorithm," vol. 5, pp. 7–16.
- [6] C. Budihartanti, C. I. Ifaru, A. Zahra, and M. H. Aenuddin, "Pengelompokan Film Pada Platform Netflix Menggunakan Metode K-Means Clustering Sebagai Rekomendasi Film," vol. 5, no. 4, pp. 1392–1402, 2024, doi: 10.47065/josh.v5i4.5482.
- [7] D. Verakandhi, P. Studi, and P. Seni, "Perubahan Preferensi Menonton Film Pada Era Media Sosial : Dampak Short Video Dan Implikasinya Terhadap," *Rekam J. Fotogr. Telev. Animasi*, vol. 20, no. 1, pp. 37–45, 2024, [Online]. Available: <https://journal.isi.ac.id/index.php/rekam/article/view/11286%0Ahttps://journal.isi.ac.id/index.php/rekam/article/download/11286/3644>
- [8] C. M. Karolina, E. Maryani, and D. W. Sjchro, "Implikasi genre film dan pemahaman penonton film tuna netra di 'Bioskop Harewos,'" *ProTVF*, vol. 4, no. 1, p. 123, 2020, doi: 10.24198/ptvf.v4i1.25035.
- [9] T. A. Sitorus and L. E. Astrianty, "Implementasi The Movie DataBase API Untuk Sistem Informasi Film Berbasis Mobile".
- [10] J. M. Informatika and S. I. Misi, "MULTI LABEL KLASIFIKASI GENRE FILM BERDASARKAN SINOPSIS MENGGUNAKAN METODE LONG SHORT-TERM MEMORY (LSTM)," vol. 8, pp. 110–119, 2025.
- [11] L. Qadrini, A. Sepperwali, and A. Aina,

- “Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial,” *J. Inov. Penelit.*, vol. 2, no. 7, pp. 1959–1966, 2021.
- [12] M. A. R. Reynaldhi and Y. Sibaroni, “Analisis Sentimen Review Film pada Twitter menggunakan Metode Klasifikasi Hybrid SVM, Naïve Bayes, dan Decision Tree,” *e-Proceeding Eng.*, vol. 8, no. 5, p. 10127, 2021.
- [13] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, “Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung,” *Inform. J. Ilmu Komput.*, vol. 18, no. 3, p. 239, 2022, doi: 10.52958/iftk.v18i3.4694.
- [14] A. H. Nasrullah, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris,” *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51, 2021, doi: 10.35329/jiik.v7i2.203.
- [15] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, and U. Firdaus, “Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank,” *Karimah Tauhid*, vol. 3, no. 2, pp. 1860–1874, 2024, doi: 10.30997/karimahtauhid.v3i2.11952.
- [16] M. Sulistiyono, Y. Pristyanto, S. Adi, and G. Gumelar, “Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi,” *Sistemasi*, vol. 10, no. 2, p. 445, 2021, doi: 10.32520/stmsi.v10i2.1303.