

PREDIKSI *RATING* FILM IMDb MENGGUNAKAN *DECISION TREE*

Rifqy Rosyidah Ilmi^{*1}, Fachrul Kurniawan², Sri Harini³

^{1,2,3}Universitas Negeri Maulana Malik Ibrahim, Malang

Email: ¹rifqyrosyidahilmi@gmail.com, ²fachrulk@ti.uin-malang.ac.id, ³sriharini@mat.uin-malang.ac.id

^{*}Penulis Korespondensi

(Naskah masuk: 14 Oktober 2022, diterima untuk diterbitkan: 25 Juli 2023)

Abstrak

Industri Film bukan hanya industri atau pusat hiburan semata melainkan menjadi pusat bisnis global. Popularitas atau kesuksesan film *box office* selalu menjadi perhatian di seluruh dunia. Data kesuksesan atau popularitas film saat ini tersedia secara online. IMDb merupakan satu dari sekian situs daring penyedia informasi yang berkaitan dengan film, acara televisi, yang meliputi sinopsis, daftar pemain, ulasan penilaian, dan tentunya pemberian rating film. Keberhasilan film dapat ditandai dengan perolehan rating yang tinggi. Prediksi rating film menjadi topik menarik untuk menilai keberhasilan film baik yang telah diproduksi maupun yang belum diproduksi. Prediksi nilai rating film dapat dimodelkan melalui *machine learning* menggunakan *decision tree*. Dari penelitian ini diperoleh kesimpulan bahwa popularitas film dan nilai vote user pada laman IMDb berpengaruh terhadap nilai rating film. Nilai akurasi penggunaan model *decision tree* pada data training, validasi dan testing berturut – turut adalah 0.7529, 0.7237 dan 0.7079.

Kata kunci: *data mining, decision tree, rating film*

IMDb MOVIE RATING PREDICTION USING DECISION TREE

Abstract

The film industry is not just an industry or entertainment but also a global business center. The popularity or success of box office movies has always been a concern around the world. Data on the success or popularity of a movie is currently available online. IMDb is one of the many online sites that provide information related to movies, television shows, which include synopsis, cast lists, rating reviews, and of course movie rating assignments. Prediction of movie ratings is an interesting topic to assess the success of films that have been produced and those that have not been produced. Prediction of movie ratings values can be modeled through machine learning using the decision tree model. From this research, it can be concluded that the popularity of the film and the value of user votes on the IMDb page have an effect on the film rating value. The accuracy values of using the descision tree model in training data, validation and testing are respectively 0.7529, 0.7237 and 0.7079.

Keywords: *data mining, decision tree, movie rating*

1. PENDAHULUAN

Popularitas atau kesuksesan film *box office* selalu menjadi perhatian di seluruh dunia. Data kesuksesan atau popularitas film yang sangat besar ini tersedia secara online (Bristi, Zaman dan Sultana, 2019). Popularitas dapat dilihat pada nilai *rating* yang saat ini dapat ditemui pada beberapa situs seperti International Movie Database (IMDb). IMDb merupakan situs daring penyedia informasi yang berkaitan dengan film, acara televisi, yang meliputi sinopsis, daftar pemain, ulasan penilaian, dan tentunya pemberian rating film yang bersangkutan.

Range nilai rating film oleh IMDb yaitu 1 (satu) sampai dengan 10 (sepuluh).

Film dibuat dengan tujuan kesuksesan yaitu merambah pasar global, menghibur penonton, mendapat rating tinggi serta memperoleh keuntungan yang besar. Sutradara yang baik dan aktor berpengalaman merupakan faktor penting dalam pembuatan film yang bagus.

Keberhasilan film dapat ditandai dengan perolehan rating yang tinggi. Prediksi rating film menjadi topik menarik untuk menilai keberhasilan film baik yang telah diproduksi maupun yang belum diproduksi.

Beberapa penelitian terkait analisis nilai rating film telah banyak dilakukan. Metode-metode yang diusulkan dapat dikategorikan berdasarkan pendekatan dalam pengumpulan data baik dengan menggunakan pemfilteran kolaboratif yang memanfaatkan data akun pengguna, memanfaatkan metadata film, atau memanfaatkan data media sosial (Abarja dan Wibowo, 2020).

Studi sebelumnya menggunakan pendekatan pemfilteran kolaboratif. Penelitian ini memanfaatkan data dari data pengguna untuk membuat prediksi sehingga peringkat yang diprediksi dibuat untuk pengguna tertentu. Pada penelitian yang dilakukan oleh (Fikir, Yaz dan Özyer, 2010) data pengguna dari IMDb dimanfaatkan dan k-Nearest Neighborhood (kNN) digunakan sebagai model. Selain itu R. Hurtado, dkk (Hurtado et al., 2019) menggunakan metode soft predict clustering dapat meningkatkan nilai prediksi dengan dataset berbasis IOT. R. Zhang, dkk (Zhang dan Mao, 2019) memanfaatkan metode markovian factorization of process (MPFP). Data pengguna juga dapat dimanfaatkan untuk memberikan rekomendasi film seperti penelitian yang dilakukan oleh (Widiyaningtyas, Hidayah dan Adji, 2021).

Selain menggunakan data dari database film, beberapa penelitian juga menggunakan media sosial sebagai sumber data. Gaikar, D, dkk (Gaikar et al., 2019) memanfaatkan data dari media sosial seperti facebook, instagram dan twitter untuk menggali data popularitas aktor, aktris, sutradara dan penulis. Penelitian tersebut menggunakan metode kNN dan menghasilkan prediksi kesuksesan sebuah film dengan rating hits, netral dan gagal. Pada penelitian yang dilakukan oleh (Mhowwala, Sulthana dan Shetty, 2020) data dari IMDb, Youtube dan wikipedia diolah menggunakan algoritma *random forest* dan *XGBoost* untuk melakukan prediksi rating film. Dari penelitian tersebut didapat akurasi 95% pada *XGBoost*.

Penelitian yang dilakukan oleh Marco Del Vecchio, dkk (Del Vecchio et al., 2021) untuk meningkatkan produktifitas perfilman Hollywood. Analisa emosi berdasarkan skrip film dapat menunjukkan prediksi kesuksesan film menggunakan k-means klustering. Kombinasi data yang tepat antara script, budget dan genre dapat menghasilkan kriteria film box office. Pada penelitian (Chakraborty, Zahidur dan Rahman, 2019), (Dixit, Hussain dan Singh, 2020) model klasifikasi dan regresi linier digunakan untuk menemukan faktor yang membuat film sukses. (Abarja dan Wibowo, 2020), (Vahidi Farashah et al., 2021) memanfaatkan *neural network* untuk melakukan prediksi rating film.

Model *machine learning* yang digunakan pada sebagian besar penelitian sebelumnya menggunakan model non-deep learning, seperti: klustering, regresi dan jaringan saraf sederhana.

2. METODE PENELITIAN

Pada penelitian ini, digunakan metode pengumpulan metadata film dan memanfaatkan model data mining *decision tree* untuk melakukan analisa dan prediksi faktor yang mempengaruhi kesuksesan suatu film yang ditandai dengan tinggi atau rendah rating film pada IMDb.

2.1 Koleksi Data

Data yang digunakan dalam penelitian ini adalah imdb movie dataset. Pada dataset terdapat 28 atribut dari 5043 film dalam waktu 100 tahun pada 66 negara yang memuat 2399 sutradara serta ribuan aktor dan aktris.

Tabel 1. Atribut Dataset

Atribut	Keterangan
movie_title	Judul Film
duration	Durasi dalam menit
director_name	Nama sutradara
director_facebook_likes	jumlah like facebook sutradara
actor_1_name	nama pemeran utama
actor_1_facebook_likes	jumlah like pada facebook aktor 1
actor_2_name	nama pemeran 2
actor_2_facebook_likes	jumlah like pada facebook pemeran 2
actor_3_name	nama pemeran 3
actor_3_facebook_likes	jumlah like pada facebook pemeran 2
num_user_for_reviews	jumlah review user
num_critic_for_reviews	jumlah critical review pada IMDb
num_voted_users	jumlah voted film pada IMDb
cast_total_facebook_likes	like facebook seluruh pemeran
movie_facebook_likes	jumlah like pada facebook film
plot_keywords	kata kunci film
facenumber_in_poster	jumlah pemeran pada poster film
color	Warna film
genres	kategori film
title_year	waktu rilis film
language	bahasa pada film
country	Negara produksi
content_rating	kualifikasi usia
aspect_ratio	aspect ratio film
movie_imdb_link	link IMDb
gross	omset film (\$)
budget	biaya produksi (\$)
imdb_score	Nilai Rating IMDb

2.2 Data Mining

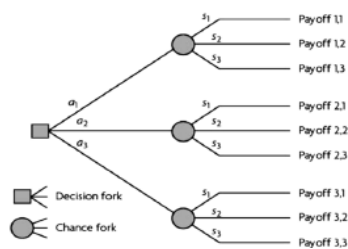
Pada data mining, data mentah yang tersimpan di dalam basis data diolah sehingga menghasilkan informasi yang berguna. Data yang disimpan hanya sebagai dokumentasi dimanfaatkan untuk

pembuatan laporan, mengontrol kebutuhan, menemukan informasi dan pengetahuan baru. Data pada database ditelusuri untuk membangun model dan mengenali pola data lain yang tidak tersimpan dalam basis data (Kasih, 2019).

Pada penelitian ini, data mining dengan model *decision tree* digunakan untuk analisis faktor yang mempengaruhi tinggi atau rendah nya nilai rating film IMDb.

2.3 Decision tree

Struktur diagram alur seperti pohon, di mana setiap simpul internal mewakili pengujian atribut. Setiap cabang mewakili hasil pengujian, dan simpul daun mewakili kelas atau distribusi kelas. Alur *decision tree* mengikuti dari simpul akar ke simpul daun yang berisi prediksi. (Kasih, 2019).



Gambar 1. Decision tree

Tahapan *Decision tree*:

1. Memilih Root Node didasarkan pada Gini Indeks dan Information Gain tertinggi.
2. Pada masing-masing iterasi algoritma ini menghitung Gini.
3. Index dan Information Gain mempertimbangkan yang setiap node tidak pernah menggunakan sebelum nya.
4. Memilih root node berdasarkan pada Gini Indeks terendah atau I.G tertinggi.
5. Kemudian memisahkan himpunan S untuk menghasilkan subset data.
6. Algoritma terus menerus berulang pada setiap subset dan memastikan bahwa atribut adalah segar dan membuat Pohon keputusan.

Information Gain dengan Nilai Entrophy Entrophy adalah formula untuk menghitung homogenitas atribut (A) dari sebuah sample data (S). Dengan formula :

$$\text{Entropy}(S) = \sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

S = Himpunan kasus dalam dataset

A = Fitur (atribut)

N = jumlah partisi atribut S

Pi = proporsi dari Si terhadap S

maka:

$$\text{Entropy}(S) = \sum_{i=1}^n p(i|s) \log_2 p(i|s)$$

Maka Gain(S,A) adalah Information Gain dari atribut A pada koleksi contoh S.

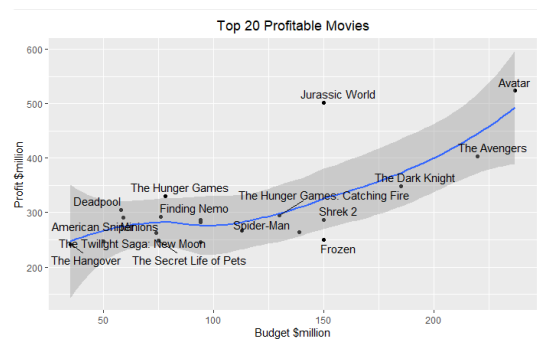
$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \text{Entropy}(S_i) \quad (2)$$

3. PENGEMBANGAN SISTEM

3.1 Representasi Pengetahuan

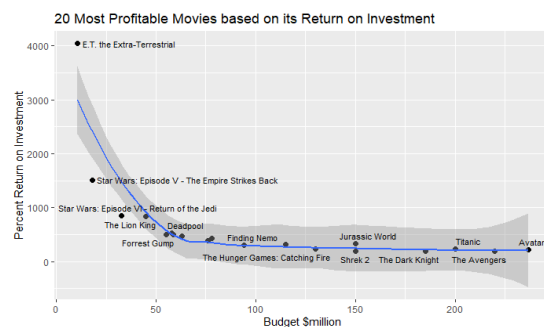
Atribut *imdb_score* pada dataset ditentukan sebagai data label (event) yaitu atribut penentu proses prediksi. Untuk meningkatkan akurasi prediksi atribut maka ditentukan pola data yang mempengaruhi event yaitu atribut yang digunakan dalam proses prediksi.

Berikut merupakan analisis dan visualisasi beberapa pola dan event data film teratas berdasarkan Laba yang diperoleh (Gross - Budget) dengan rentang tahun rilis 2000 - 2006.



Gambar 2. Top 20 Film Laba tertinggi

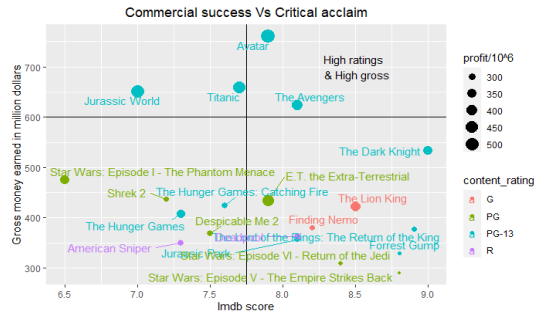
Tren kenaikan hampir linier, profit (gross) meningkat seiring dengan kenaikan anggaran (budget). Film teratas berdasarkan Persentase Pengembalian Investasi ((laba / anggaran) * 100).



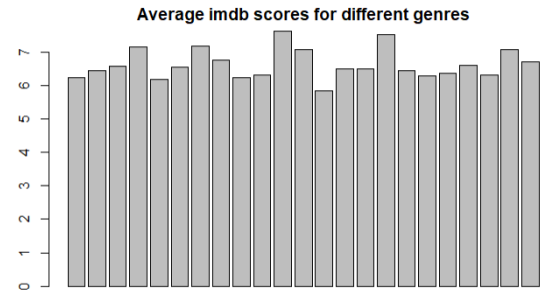
Gambar 3. Top 20 Film ROI tertinggi

Keuntungan yang diperoleh sebuah film tidak memberikan gambaran yang jelas tentang kesuksesan moneternya, perbandingan nilai Pengembalian Investasi (ROI) terhadap anggaran akan memberikan hasil yang lebih baik. Seperti yang dihipotesiskan, ROI tinggi untuk Film Anggaran Rendah dan menurun seiring dengan peningkatan anggaran film.

Gambar 3 menunjukkan korelasi antara keuntungan komersial dan nilai rating IMDb tidak banyak karena sebagian besar film yang mendapat pujian kritis tidak menghasilkan banyak keuntungan secara komersial.



Gambar 4. Korelasi Keuntungan Komersial dan IMDb



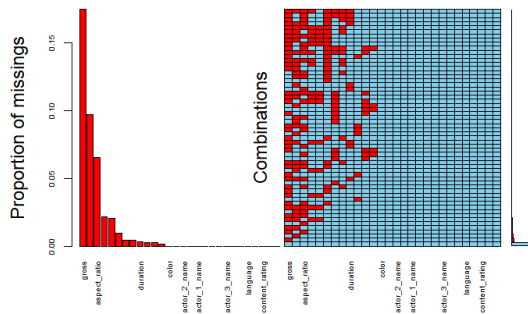
Gambar 7. Rata – Rata nilai IMDb sesuai Genre

3.2 Decision tree

3.2.1 Data Cleaning

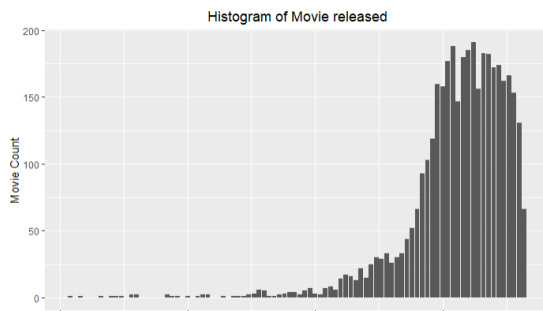
Proses cleaning dilakukan agar data sesuai dengan kebutuhan metode prediksi yang digunakan. Proses cleaning dilakukan dengan tujuan:

- Menghapus data yang memiliki *missing value*.



Gambar 5. Atribut dengan NA

- Menghapus atribut yang tidak relevan diantaranya yaitu *aspect_ratio*, *color*, *language*.
- Analisis pencilaan data : film yang dirilis sebelum tahun 1980. Data yang digunakan adalah data film yang rilis setelah tahun 1980.



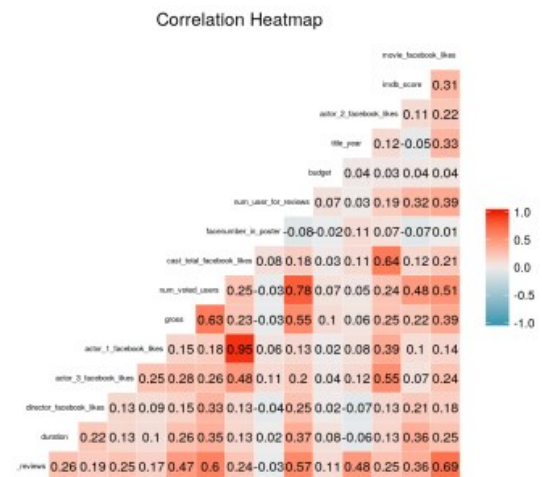
Gambar 6. Analisis Pencilaan Data

- Pengelompokan atribut.
Atribut *country* dikelompokkan menjadi 3 yaitu USA, UK dan others.
Atribut *content_rating* dikelompokkan menjadi 5 yaitu G, NC-17, PG, PG-13 dan R.
- Mendeteksi perbedaan data : misal pada jumlah like facebook 0, data tersebut masuk pada data error dan diberi nilai rata- rata.
- Menyeleksi subset atribut : split data genre yang terdapat tanda |. Dihitung nilai rata- rata IMDb sesuai genre.

Nilai IMDb cenderung konstan sehingga atribut genre tidak dijadikan sebagai predictor.

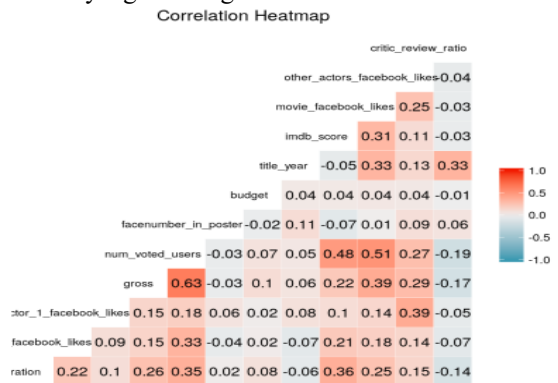
3.2.2 Data Preprocessing

Pada preprocessing data dilakukan plotting korelasi antar atribut.



Gambar 8. Korelasi Atribut

Nilai korelasi tertinggi yaitu 0,95 yang menunjukkan atribut pada data dapat dianalisis untuk dilakukan prediksi. Atribut yang saling berkorelasi dimodifikasi sehingga diperoleh atribut relevan yang akan digunakan untuk analisis data.



Gambar 9. Atribut untuk analisis

Prediksi film baik atau buruk menggunakan nilai rating yang telah tersedia. Guna menentukan prediksi sebuah film tersebut baik atau buruk maka

dilakukan pengelompokan skor menjadi 4 kelompok yaitu : kurang dari 4, 4 - 6, 6 -8 dan 8 - 10, yang masing-masing mewakili buruk, OK, baik dan sangat baik. Kelompok tersebut diinisasi pada atribut `binned_score`.

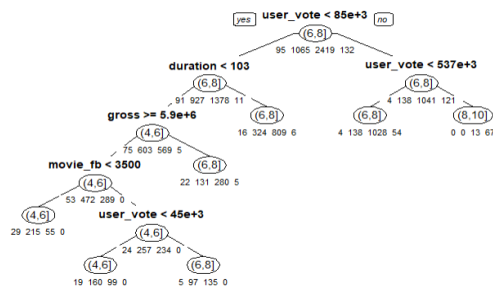
Setelah dilakukan proses cleaning maka dapat diperoleh dataset bersih atau atribut bersih.

3.3 Implementasi

Setelah dilakukan proses data cleaning dan preprosesing maka dilakukan proses data mining menggunakan model *decision tree*.

Menggunakan *decision tree* sebagai klasifikasi, dihasilkan representasi grafis dari faktor yang menunjang tinggi atau rendahnya nilai rating film berdasarkan atribut seperti aktor atau sutradara favorit, anggaran film, popularitas film. *Decision tree* dimulai dengan satu root yang kemudian bercabang menjadi sejumlah solusi. Kriteria yang digunakan untuk klasifikasi adalah *gain ratio*, karena dapat menentukan relevansi atribut dengan baik.

Decision tree ditunjukkan pada gambar berikut:



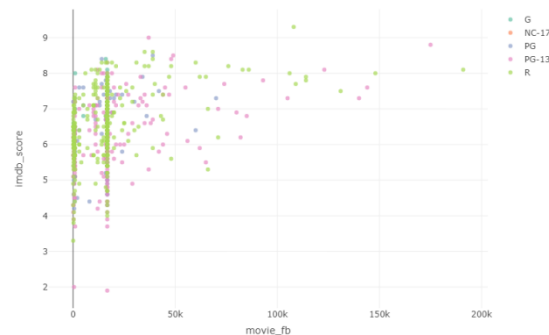
Gambar 10. Decision Tree

Dari Gambar 10 maka decision tree memiliki *rules* sebagai berikut.

Rule	Kondisi
R1	Jika (<code>user_vote</code> ≥ 537000) maka kelompok (8,10)
R2	Jika (<code>85000</code> \leq <code>user_vote</code> $<$ <code>551000</code>) maka kelompok = (6,8).
R3	Jika (<code>user_vote</code> $<$ <code>85000</code>) dan (<code>duration</code> ≥ 106) maka kelompok = (6,8).
R4	Jika (<code>user_vote</code> $<$ <code>85000</code>), (<code>duration</code> $<$ <code>106</code>) dan (<code>gross</code> $<$ <code>5900000</code>) maka kelompok = (6,8).
R5	Jika (<code>user_vote</code> $<$ <code>850000</code>), (<code>duration</code> $<$ <code>106</code>), (<code>gross</code> ≥ 5900000) dan (<code>movie_fb</code> $<$ <code>4500</code>) maka kelompok = (4,6).
R6	Jika (<code>user_vote</code> $<$ <code>85000</code>), (<code>duration</code> $<$ <code>106</code>), (<code>gross</code> ≥ 5900000) dan (<code>movie_fb</code> ≥ 4500) maka kelompok = (6,8).

Film dengan banyak vote pada situs IMDb cenderung memiliki skor yang lebih tinggi dan masuk ke dalam kelompok sangat baik, hal tersebut dikarenakan film populer akan memiliki banyak penggemar untuk melakukan voting dan memberikan skor tinggi.

Korelasi antara jumlah like facebook atau kepopuleran film dan `imdb_score` dapat direpresentasikan pada gambar berikut.



Gambar 14. Visualisasi Facebook Like

Jumlah like di Facebook yang tinggi menyebabkan rating film cenderung tinggi. Rating film dengan like di Facebook yang rendah bervariasi dalam rentang yang luas.

Jika sebuah film memiliki vote yang lebih sedikit, film tersebut masih bisa menjadi film yang bagus jika durasinya lebih lama sesuai dengan R4. Film yang menghasilkan lebih sedikit keuntungan masuk ke dalam kategori baik, tetapi masuk dalam kelompok OK jika film menghasilkan lebih banyak keuntungan.

4. PENGUJIAN

Pengujian sistem dilakukan untuk mengetahui akurasi penerapan metode *decision tree* pada proses training, validasi dan testing. Pengujian dilakukan menggunakan confusion matrix. Data dibagi menjadi training, validasi dan testing dengan perbandingan 6:2:2.

Hasil akurasi penerapan *decision tree* pada data training adalah sebagai berikut:

Confusion Matrix and Statistics				
Reference				
Prediction	(0,4]	(4,6]	(6,8]	(8,10]
(0,4]	0	0	0	0
(4,6]	33	295	94	0
(6,8]	30	355	1342	34
(8,10]	0	0	4	39
Overall Statistics				
Accuracy : 0.7529				
95% CI : (0.7345, 0.7707)				
No Information Rate : 0.6469				
P-Value [Acc > NIR] : < 2.2e-16				
Kappa : 0.4284				
McNemar's Test P-Value : NA				
Statistics by Class:				
	Class: (0,4]	Class: (4,6]	Class: (6,8]	Class: (8,10]
Sensitivity	0.0000	0.4538	0.9319	0.53425
Specificity	1.0000	0.9134	0.4669	0.99814
Pos Pred Value	NaN	0.6991	0.7521	0.90698
Neg Pred Value	0.9717	0.8032	0.7892	0.98443
Prevalence	0.0283	0.2920	0.6469	0.03279
Detection Rate	0.0000	0.1325	0.6029	0.01752
Detection Prevalence	0.0000	0.1896	0.7911	0.01932
Balanced Accuracy	0.5000	0.6866	0.6994	0.76619

Gambar 15. Hasil Akurasi Data Training

Berdasarkan data diatas, pada data training diperoleh prediksi kelompok (8,10) atau film dengan rating sangat baik sebanyak 39 film dan dapat diprediksi dengan benar, kelompok (6,8) rating baik sebanyak 1342 film dan diprediksi dengan benar. Secara keseluruhan akurasi prediksi menggunakan

model *decision tree* yaitu sebesar 0,7529 atau 75,29% pada data training.

Hasil akurasi penerapan *decision tree* pada validasi adalah sebagai berikut:

Confusion Matrix and Statistics

	Reference			
Prediction	(0,4]	(4,6]	(6,8]	(8,10]
(0,4]	0	0	0	0
(4,6]	13	61	41	0
(6,8]	9	121	465	16
(8,10]	0	0	5	11

Overall Statistics

Accuracy : 0.7237
 95% CI : (0.69, 0.7556)
 No Information Rate : 0.6887
 P-Value [Acc > NIR] : 0.02076
 Kappa : 0.299
 McNemar's Test P-Value : NA

Statistics by Class:

	Class: (0,4]	Class: (4,6]	Class: (6,8]	Class: (8,10]
Sensitivity	0.00000	0.33516	0.9100	0.40741
Specificity	1.00000	0.90357	0.3680	0.99301
Pos Pred Value	NaN	0.53043	0.7610	0.68750
Neg Pred Value	0.97035	0.80702	0.6489	0.97796
Prevalence	0.02965	0.24528	0.6887	0.03639
Detection Rate	0.00000	0.08221	0.6267	0.01482
Detection Prevalence	0.00000	0.15499	0.8235	0.02156
Balanced Accuracy	0.50000	0.61937	0.6390	0.70021

Gambar 16. Hasil Akurasi Data Validasi

Berdasarkan data diatas, pada data validasi diperoleh prediksi kelompok (8,10) atau film dengan rating sangat baik sebanyak 11 film dan dapat diprediksi dengan benar, kelompok (6,8) rating baik sebanyak 465 film dan diprediksi dengan benar. Secara keseluruhan akurasi prediksi menggunakan model *decision tree* yaitu sebesar 0,7237 atau 72,37% pada data validasi.

Hasil akurasi penerapan *decision tree* pada data testing adalah sebagai berikut:

Confusion Matrix and Statistics

	Reference			
Prediction	(0,4]	(4,6]	(6,8]	(8,10]
(0,4]	0	0	0	0
(4,6]	4	82	35	0
(6,8]	6	151	429	17
(8,10]	0	0	4	15

Overall Statistics

Accuracy : 0.7079
 95% CI : (0.6738, 0.7404)
 No Information Rate : 0.6299
 P-Value [Acc > NIR] : 4.486e-06
 Kappa : 0.3311
 McNemar's Test P-Value : NA

Statistics by Class:

	Class: (0,4]	Class: (4,6]	Class: (6,8]	Class: (8,10]
Sensitivity	0.00000	0.3519	0.9167	0.46875
Specificity	1.00000	0.9235	0.3673	0.99437
Pos Pred Value	NaN	0.6777	0.7114	0.78947
Neg Pred Value	0.98654	0.7572	0.7214	0.97652
Prevalence	0.01346	0.3136	0.6299	0.04307
Detection Rate	0.00000	0.1104	0.5774	0.02019
Detection Prevalence	0.00000	0.1629	0.8116	0.02557
Balanced Accuracy	0.50000	0.6377	0.6420	0.73156

Gambar 17. Hasil Akurasi Data Testing

Kelompok (8,10) atau film dengan rating sangat baik sebanyak 15 film dan dapat diprediksi dengan benar, kelompok (6,8) rating baik sebanyak 429 film dan diprediksi dengan benar. Secara keseluruhan akurasi prediksi menggunakan model *decision tree* yaitu sebesar 0,7079 atau 70,79% pada data testing.

5. KESIMPULAN

Hasil penelitian menunjukkan *decision tree* mampu melakukan prediksi nilai rating film IMDb dan menunjukkan faktor yang mempengaruhi tinggi atau rendah nilai rating film.

Jumlah like di Facebook yang tinggi menunjukkan film populer dan memiliki banyak

penggemar untuk melakukan voting dan memberikan skor tinggi.

Nilai akurasi *decision tree* pada data training, validasi dan testing berturut – turut adalah 0.7529, 0.7237 dan 0,7079.

DAFTAR PUSTAKA

- ABARJA, R.A. DAN WIBOWO, A., 2020. Movie rating prediction using convolutional neural network based on historical values. *International Journal of Emerging Trends in Engineering Research*, 8(5), hal.2156–2164.
<https://doi.org/10.30534/ijeter/2020/10985> 2020.
- BRISTI, W.R., ZAMAN, Z. DAN SULTANA, N., 2019. Predicting IMDb Rating of Movies by Machine Learning Techniques. *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*, (May 2020).
<https://doi.org/10.1109/ICCCNT45670.2019.8944604>.
- CHAKRABORTY, P., ZAHIDUR, M. DAN RAHMAN, S., 2019. Movie Success Prediction using Historical and Current Data Mining. *International Journal of Computer Applications*, 178(47), hal.1–5.
<https://doi.org/10.5120/ijca2019919415>.
- DIXIT, P., HUSSAIN, S. DAN SINGH, G., 2020. Predicting the IMDB rating by using EDA and machine learning Algorithms. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, (August), hal.441–446.
<https://doi.org/10.32628/cseit206481>.
- FIKIR, O.B., YAZ, I.O. DAN ÖZYER, T., 2010. A movie rating prediction algorithm with collaborative filtering. *Proceedings - 2010 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2010*, hal.321–325.
<https://doi.org/10.1109/ASONAM.2010.6>.
- GAIKAR, D., SOLANKI, R., SHINDE, H., PHAPALE, P. DAN PANDEY, I., 2019. Movie Success Prediction Using Popularity Factor from Social Media. *International Research Journal of Engineering and Technology*, (daring) 6(4), hal.5185–5190.
 Tersedia pada: <www.irjet.net>.
- HURTADO, R., BOBADILLA, J., BOJORQUE, R., ORTEGA, F. DAN LI, X., 2019. A new recommendation approach based on probabilistic soft clustering methods: A scientific documentation case study. *IEEE Access*, 7, hal.7522–7534.

- <https://doi.org/10.1109/ACCESS.2018.2890079>.
- KASIH, P., 2019. Pemodelan Data Mining Decision Tree Dengan Kelompokifikasi Error Untuk Seleksi Calon Anggota Tim Paduan Suara. *Innovation in Research of Informatics (INNOVATICS)*, 1(2), hal.63–69.
<https://doi.org/10.37058/innovatics.v1i2.918>.
- MHOWWALA, Z., SULTHANA, A.R. DAN SHETTY, S.D., 2020. Movie rating prediction using ensemble learning algorithms. *International Journal of Advanced Computer Science and Applications*, 11(8), hal.383–388.
<https://doi.org/10.14569/IJACSA.2020.0110849>.
- VAHIDI FARASHAH, M., ETEBARIAN, A., AZMI, R. DAN EBRAHIMZADEH DASTJERDI, R., 2021. A hybrid recommender system based-on link prediction for movie baskets analysis. *Journal of Big Data*, 8(1).
<https://doi.org/10.1186/s40537-021-00422-0>.
- DEL VECCHIO, M., KHARLAMOV, A., PARRY, G. DAN POGREBNA, G., 2021. Improving productivity in Hollywood with data science: Using emotional arcs of movies to drive product and service innovation in entertainment industries. *Journal of the Operational Research Society*, 72(5), hal.1110–1137.
<https://doi.org/10.1080/01605682.2019.1705194>.
- WIDIYANINGTYAS, T., HIDAYAH, I. DAN ADJI, T.B., 2021. User profile correlation-based similarity (UPCSim) algorithm in movie recommendation system. *Journal of Big Data*, (daring) 8(1).
<https://doi.org/10.1186/s40537-021-00425-x>.
- ZHANG, R. DAN MAO, Y., 2019. Movie Recommendation via Markovian Factorization of Matrix Processes. *IEEE Access*, 7, hal.13189–13199.
<https://doi.org/10.1109/ACCESS.2019.2892289>.

Halaman ini sengaja dikosongkan