



Artikel

Aplikasi Prediksi *Rating* Film dengan Perbandingan Metode Naïve Bayes dan KNN Berbasis *Website* Menggunakan *Framework* Codeigniter

Andrew Ryan Yosafat¹, Yusuf Kurnia²

^{1,2} Universitas Buddhi Dharma, Teknik Informatika, Banten, Indonesia

JEJAK PENGIRIMAN

Received: 28 Agustus 2019

Final Revision: 25 September 2019

Available Online: 30 September 2019

KATA KUNCI

Prediksi, Rating Film, Naïve Bayes, KNN, Data mining

KORESPONDENSI

Telepon: 089503263415

E-mail: andrewmishima@gmail.com

A B S T R A K

Pada saat ini industri perfilman di dunia sudah sangat hebat, dikarenakan perkembangan teknologi yang sangat pesat sehingga memberi pengaruh yang besar terhadap industri perfilman. Berkat perkembangan teknologi inilah yang membuat film-film yang dibuat menjadi semakin menarik, bahkan banyak munculnya kreator-kektor baru yang belajar membuat film dengan *budget* yang kecil mulai dari membuat film pendek maupun film cerita (*feature film*). Film yang berhasil cenderung memiliki *rating* film tinggi, sungguh menarik bukan jika dapat mengetahui keberhasilan suatu film yang telah dibuat atau belum dibuat dengan memprediksi seberapa tinggi *rating* film tersebut. Untuk memprediksi *rating* film digunakanlah metode *data mining* yaitu Naïve Bayes dan KNN, Naïve Bayes dan KNN adalah metode pengelompokan statistik yang dipakai untuk memprediksi probabilitas anggota suatu class. Prediksi menggunakan faktor internal dan eksternal yang mempengaruhi keberhasilan film sebagai variabel *independent* dan *rating* film sebagai variabel *dependent*. Dari sistem yang dibuat dapat diketahui *rating* film sebuah film dan faktor-faktor yang mempengaruhi keberhasilan sebuah film. Dengan sistem ini diharapkan dapat mengefektifkan dan mengefisiensikan produksi sebuah film.

PENDAHULUAN

Industri perfilman adalah salah satu jenis industri yang sedang naik daun dan setiap tahun semakin meningkat baik dari munculnya industri film baru maupun film yang diproduksi oleh industri.

Film adalah sarana baru yang dipergunakan untuk menyebarkan suatu hiburan yang telah menjadi kebiasaan terdahulu, dan menyajikan cerita peristiwa, musik, drama, lawak, dan sajian teknis lainnya kepada masyarakat umum. [1]

Media komunikasi baik media cetak, media elektronik, ataupun media internet

juga sering menyajikan hasil *review* dari sebuah film, dalam sebuah *review* film biasanya disertakan *rating* film. *Rating* film adalah klasifikasi atau peringkat dari suatu film berdasarkan kualitas, standar, atau kinerja yang dinilai oleh penikmatnya.[2]

Untuk mendapatkan keuntungan yang besar tentu saja film harus memiliki *rating* film yang tinggi, karena sangat berpengaruh dengan minat masyarakat untuk menonton sebuah film.

Rating dari sebuah film diperoleh dari *feedback* masyarakat setelah menonton film. Untuk mengetahui *rating* dari sebuah film sebelum film tersebut dibuat atau didistribusikan dapat digunakan metode *data mining* untuk melakukan proses prediksi.

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki.[3]

Informasi yang didapatkan dari hasil *data mining* bisa digunakan untuk meningkatkan pendapatan atau mengurangi biaya produksi.

I. METODE

Untuk memperoleh hasil terbaik dilakukan perbandingan antara 2 metode *data mining* yaitu Naïve Bayes dan KNN.

Naïve Bayes adalah sebuah pengelompokan statistik yang bisa di dipakai untuk memprediksi probabilitas anggota suatu *class*. Naïve Bayes juga mempunyai akurasi dan kecepatan yang sangat kuat ketika diaplikasikan pada *database* dengan *big data*.

Berikut rumus Naïve Bayes:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Keterangan :

- X : Data dengan *class* yang belum diketahui
- Y : Hipotesis data yaitu suatu *class* spesifik

P(Y|X): Probabilitas hipotesis berdasar kondisi X (*posteriori probability*)

P(Y) : Probabilitas hipotesis Y (*prior probability*)

P(X|Y): Probabilitas X saat kondisi hipotesis Y

P (X) : Probabilitas X

K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. [4]

Berikut rumus KNN:

$$distance = \sqrt{\sum_{i=1}^n (X_{training}^i - X_{testing})^2}$$

- $X_{training}^i$: data training ke-i,
- $X_{testing}$: data testing,
- I : record (baris) ke-i dari tabel,
- n : jumlah data training.

II. HASIL

Data yang digunakan dalam penelitian ini adalah data sekunder, yang diperoleh dari website kaggle.com, website tersebut adalah forum peneliti *data mining* dan *machine learning*.

Dataset yang diperoleh berisikan 28 variabel, yaitu:

Tabel 1. Variabel *Dataset*

Jenis Variabel	Nama Variabel
X1	color
X2	director_name
X3	num_critc_for_reviews
X4	duration
X5	director_facebook_likes
X6	actor_3_facebook_likes
X7	actor_2_name
X8	actor_1_facebook_likes
X9	gross
X10	genres
X11	actor_1_name
X12	movie_title
X13	num_voted_users

X14	cast_total_facebook_likes
X15	actor_3_name
X16	facenumber_in_poster
X17	plot_keywords
X18	movie_imdb_link
X19	num_user_for_reviews
X20	language
X21	country
X22	content_rating
X23	Budget
X24	title_year
X25	actor_2_facebook_likes
Y	imdb_score
X26	aspect_ratio
X27	movie_facebook_likes

Untuk meningkatkan akurasi prediksi variabel yang digunakan adalah 13 variabel, yaitu:

Tabel 2. Variabel yang digunakan

No.	Variabel	Alasan
1	actor_1_facebook_likes	Menunjukkan popularitas aktor utama.
2	actor_2_facebook_likes	Menunjukkan popularitas aktor kedua.
3	actor_3_facebook_likes	Menunjukkan popularitas aktor ketiga.
4	aspect_ratio	Ukuran layar bioskop bervariasi sehingga aspect ratio mempengaruhi kenyamanan pada saat film diputar.
5	budget	Biaya produksi yang lebih besar meningkatkan kualitas film.
6	cast_total_facebook_likes	Menunjukkan popularitas <i>casting</i> film.
7	content_rating	Usia penonton bervariasi sehingga sebagian film hanya dapat ditonton oleh umur tertentu.

8	country	Negara yang lebih terkenal dengan perfilmanya meningkatkan jumlah peminat.
9	director_facebook_likes	Menunjukkan popularitas sutradara.
10	duration	Durasi sebuah film mempengaruhi minat penonton.
11	language	Penonton lebih nyaman dalam menonton film dengan bahasa tertentu.
12	movie_facebook_likes	Menunjukkan popularitas film.
13	imdb_score	Rating film IMDB yang digunakan sebagai acuan prediksi.

Setelah menentukan variabel yang akan digunakan, dilakukan *data cleansing* agar data sesuai dengan kebutuhan metode prediksi yang digunakan. Proses yang dilakukan dalam *data cleansing*, yaitu:

- a. Menghapus data yang memiliki *missing value*, data yang valid berjumlah 1820 data.
- b. Melakukan transformasi data numerik menjadi kategorikal pada variabel:
 - 1) duration
 - 2) director_facebook_likes
 - 3) actor_1_facebook_likes
 - 4) actor_2_facebook_likes
 - 5) actor_3_facebook_likes
 - 6) cast_total_facebook_likes
 - 7) movie_facebook_likes
 - 8) budget

dengan ketentuan:

- 1) Short/Low <=QUARTILE 1
- 2) Moderate/Medium >=QUARTILE 1 dan <=QUARTILE 3
- 3) Long/High >=QUARTILE 3

Tabel 3. Data Numerik

duratio n	director_facebook_l ikes	actor_1_facebook_l ikes
147	94	21000
122	681	998
110	282	11000
123	452	10000
106	4000	19000

Tabel 4. Data Kategorikal

duratio n	director_facebook_ likes	actor_1_facebook_ likes
Long	Medium	High
Long	High	Medium
Moderate	Medium	Medium
Long	High	Medium
Moderate	High	High

Setelah dilakukan proses *data cleansing*, *dataset* diuji dengan algoritma Naïve Bayes dan KNN pada aplikasi RapidMiner.

RapidMiner merupakan perangkat lunak untuk pengolahan data, dengan menggunakan prinsip dan algoritma datamining Rapidminer mengekstrak pola-

a. *Class Probability*

imdb score	P(Y)
High	11 0,220
Medium	21 0,420
Low	18 0,360

Gambar 1: *Class Probability*

b. *Conditional Probability*

Duration	Laplace C. (H,M,L)			Laplace C. (H,M,L)		
	High	Medium	Low	P(high)	P(med)	P(low)
Long	6	9	1	0,545	0,429	0,056
Moderate	4	10	6	0,364	0,476	0,333
Short	1	2	11	0,091	0,095	0,611

director_facebook_likes	Laplace C. (H,M,L)			Laplace C. (H,M,L)		
	High	Medium	Low	P(high)	P(med)	P(low)
High	4	10	4	0,364	0,476	0,222
Medium	3	9	5	0,273	0,429	0,278
Low	4	2	9	0,364	0,095	0,500

actor_1_facebook_likes	Laplace C. (H,M,L)			Laplace C. (H,M,L)		
	High	Medium	Low	P(high)	P(med)	P(low)
High	5	10	1	0,455	0,476	0,056
Medium	3	9	11	0,273	0,429	0,611
Low	3	2	6	0,273	0,095	0,333

actor_2_facebook_likes	Laplace C. (H,M,L)			Laplace C. (H,M)		
	High	Medium	Low	P(high)	P(med)	P(low)
High	5	11	1	0,455	0,524	0,056
Medium	1	7	8	0,091	0,333	0,444
Low	5	3	9	0,455	0,143	0,500

Language	Laplace C. (H,M,L)			Laplace C. (H,M,L)		
	High	Medium	Low	P(high)	P(med)	P(low)
English	10	20	18	0,909	0,952	1,000
Japanese	0	1	0	0,000	0,048	0,000
Portuguese	1	0	0	0,091	0,000	0,000

Country	Laplace C. (H,M,L)			Laplace C. (H,M,L)		
	High	Medium	Low	P(high)	P(med)	P(low)
USA	7	19	17	0,636	0,905	0,944
UK	1	0	0	0,091	0,000	0,000
Australia	1	0	0	0,091	0,000	0,000
China	1	0	0	0,091	0,000	0,000
France	0	1	0	0,000	0,048	0,000
Norway	0	0	1	0,000	0,000	0,056
Japan	0	1	0	0,000	0,048	0,000
Brazil	1	0	0	0,091	0,000	0,000

content_rating	Laplace C. (H,M)			Laplace C. (H,M)		
	High	Medium	Low	P(high)	P(med)	P(low)
R	3	5	9	0,273	0,238	0,500
PG	2	4	2	0,182	0,190	0,111
PG-13	6	12	6	0,545	0,571	0,333
Not Rated	0	0	1	0,000	0,000	0,056

pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database.[5] Sehingga diperoleh hasil sebagai berikut:

Tabel 5. Hasil perbandingan algoritma

	Naïve Bayes	KNN
Akurasi	53,99%	51,26%

III. PEMBAHASAN

Untuk mengetahui apakah rumus dalam aplikasi sudah benar dilakukan pengujian pada hasil prediksi aplikasi, hasil prediksi RapidMiner dan hasil prediksi perhitungan manual yang dihitung dalam MS. Excel. Data yang digunakan dalam pengujian adalah 50 data yang dipilih secara acak.

1. Perhitungan Manual

Perhitungan yang dilakukan secara manual dengan bantuan MS. Excel dalam perhitungannya. Langkah-langkah dalam melakukan perhitungan Naïve Bayes adalah sebagai berikut:

actor_3_facebook_likes							budget						
	High	Medium	Low	P(high)	P(med)	P(low)		High	Medium	Low	P(high)	P(med)	P(low)
High	3	7	2	0,273	0,333	0,111	High	7	12	2	0,636	0,571	0,111
Medium	4	9	7	0,364	0,429	0,389	Medium	1	6	5	0,091	0,286	0,278
Low	4	5	9	0,364	0,238	0,500	Low	3	3	11	0,273	0,143	0,611
cast_total_facebook_likes							aspect_ratio						
	High	Medium	Low	P(high)	P(med)	P(low)		High	Medium	Low	P(high)	P(med)	P(low)
High	5	12	1	0,455	0,571	0,056	2.35	10	14	8	0,909	0,667	0,444
Medium	2	7	9	0,182	0,333	0,500	2.2	0	1	0	0,000	0,048	0,000
Low	4	2	8	0,364	0,095	0,444	2.0	0	1	0	0,000	0,048	0,000
movie_facebook_likes							Laplace C. (H,L)						
	High	Medium	Low	P(high)	P(med)	P(low)		High	Medium	Low	P(high)	P(med)	P(low)
High	8	14	4	0,727	0,667	0,222	1.85	1	3	7	0,091	0,143	0,389
Medium	0	6	9	0,000	0,286	0,500	1.78	0	1	2	0,000	0,048	0,111
Low	3	1	5	0,273	0,048	0,278	1.66	0	1	0	0,000	0,048	0,000
							1.33	0	0	1	0,000	0,000	0,056

Gambar 2: Conditional Probability

c. Hasil Prediksi

	A	B	C	D	E	F	G	H	I	J
1	High Class Prob	Medium Class Prob	Low Class Prob	Normalization	High Class Prob	Medium Class Prob	Low Class Prob	Prob Value	Prediction	Imdb_score
2	0,000112888	0,0004583285	0,0000000004	0,0005696177	0,1953745765	0,8046247816	0,0000006420	1	Medium	High
3	0,0000949665	0,0000148808	0,0000010426	0,0001108898	0,8564035811	0,1341944029	0,0094020160	1	High	High
4	0,0000593540	0,0003273775	0,0000001086	0,0003868402	0,1534329762	0,8462862802	0,0002807437	1	Medium	Medium
5	0,0000474832	0,0000937490	0,0000009123	0,0001421445	0,3340489844	0,6595331485	0,0064178671	1	Medium	Medium
6	0,0000006281	0,0000000000	0,0000000000	0,0000006281	1,0000000000	0,0000000000	0,0000000000	1	High	High
7	0,0000094966	0,0002187477	0,0000008109	0,0002290553	0,0414600659	0,9549997322	0,0035402019	1	Medium	Low
8	0,0000024731	0,0000363753	0,0000000006	0,0000388490	0,0636588991	0,9363246290	0,0000164719	1	Medium	Medium
9	0,0000593540	0,0003273775	0,0000001086	0,0003868402	0,1534329762	0,8462862802	0,0002807437	1	Medium	Medium
10	0,0000047483	0,0002187477	0,0000060818	0,0002295778	0,0206828496	0,9528260697	0,0264910807	1	Medium	Low
11	0,0000082436	0,0000075445	0,0000000024	0,0000157905	0,5220606985	0,4777864631	0,0001528383	1	High	High
12	0,0000000000	0,0000000067	0,0000692751	0,0000692818	0,0000000000	0,0000974088	0,9999025912	1	Low	Low
13	0,0000076312	0,0000019070	0,0000228066	0,0000305441	0,2498428025	0,0034799279	0,7466772696	1	Low	High
14	0,0000075370	0,0000064300	0,0000035987	0,0000175657	0,4290760074	0,3660527964	0,2048711963	1	High	High
15	0,0000000000	0,0000007440	0,0000508017	0,0000515458	0,0000000000	0,0144345456	0,9855654544	1	Low	Low
16	0,0000000000	0,0000019070	0,0000000000	0,0000190970	0,0000000000	1,0000000000	0,0000000000	1	Medium	Medium
17	0,0000000000	0,0000052083	0,0000000000	0,0000052083	0,0000000000	1,0000000000	0,0000000000	1	Medium	Medium
18	0,0001483851	0,0005092539	0,0000000003	0,0006576393	0,2256329442	0,7743666110	0,0000004448	1	Medium	Medium
19	0,0001128888	0,0004583285	0,0000000004	0,0005696177	0,1953745765	0,8046247816	0,0000006420	1	Medium	Medium
20	0,0000105989	0,0000000000	0,0000000000	0,0000105989	1,0000000000	0,0000000000	0,0000000000	1	High	High
21	0,0000127187	0,0000000000	0,0000000000	0,0000127187	1,0000000000	0,0000000000	0,0000000000	1	High	High
22	0,0000000000	0,00000000394	0,0000000516	0,0000000910	0,0000000000	0,4326677973	0,5673322027	1	Low	Medium
23	0,0000087932	0,0001543194	0,0000000164	0,0001631289	0,0539033145	0,9459962628	0,0001004227	1	Medium	Medium
24	0,0000002261	0,0000010802	0,0000273679	0,0000286743	0,0078854845	0,0376726270	0,9544418886	1	Low	Low
25	0,0000000000	0,0000021701	0,0001655760	0,0001677461	0,0000000000	0,0129369068	0,9870630932	1	Low	Low
26	0,0000000000	0,0000026524	0,0000072574	0,0000099098	0,0000000000	0,2676518490	0,7323481510	1	Low	Low
27	0,0000009793	0,0000004220	0,0041911432	0,0041925445	0,0002335912	0,0001006470	0,9996657618	1	Low	Low
25	0,0000000000	0,0000021701	0,0001655760	0,0001677461	0,0000000000	0,0129369068	0,9870630932	1	Low	Low
26	0,0000000000	0,0000026524	0,0000072574	0,0000099098	0,0000000000	0,2676518490	0,7323481510	1	Low	Low
27	0,0000009793	0,0000004220	0,0041911432	0,0041925445	0,0002335912	0,0001006470	0,9996657618	1	Low	Low
28	0,0000000000	0,0000000011	0,0000000000	0,0000000011	0,0000000000	1,0000000000	0,0000000000	1	Medium	Medium
29	0,0000012719	0,0000000011	0,0001451478	0,0001464208	0,0086864191	0,0000074684	0,9913061125	1	Low	Low
30	0,0000076312	0,0000000049	0,0000131953	0,0000208314	0,3663330246	0,0002362249	0,6334307505	1	Low	High
31	0,0000000000	0,0000000021	0,0000362870	0,0000362891	0,0000000000	0,0000581155	0,9999418845	1	Low	Low
32	0,0000000000	0,0000219296	0,0000000000	0,0000219296	0,0000000000	1,0000000000	0,0000000000	1	Medium	Medium
33	0,0000000000	0,0000210457	0,0000000018	0,0000210475	0,0000000000	0,9999144903	0,0000855097	1	Medium	Medium
34	0,0000003392	0,0000455724	0,0000228066	0,0000687182	0,0049356041	0,6631784026	0,3318859933	1	Medium	Medium
35	0,0000002120	0,0000347219	0,0000000384	0,0000349722	0,0060613433	0,9928407864	0,0010978703	1	Medium	Low
36	0,0000094966	0,0002187477	0,0000008109	0,0002290553	0,0414600659	0,9549997322	0,0035402019	1	Medium	High
37	0,0000000000	0,0000454691	0,0000043984	0,0000498675	0,0000000000	0,9117979271	0,0882020729	1	Medium	Medium
38	0,0000000000	0,0000584603	0,0000000648	0,0000585251	0,0000000000	0,9988929270	0,0011070730	1	Medium	Medium
39	0,0000000000	0,0000195310	0,0000068420	0,0000263730	0,0000000000	0,7405688899	0,2594311101	1	Medium	Medium
40	0,0000105989	0,0001060946	0,0000000011	0,0001166946	0,0908262704	0,9091643290	0,0000949006	1	Medium	Medium
41	0,0000000000	0,0000003163	0,0000461834	0,0000464997	0,0000000000	0,0068031235	0,9931968765	1	Low	Low
42	0,0000000000	0,0000000000	0,0000009487	0,0000009487	0,0000000000	0,0000000000	1,0000000000	1	Low	Low
43	0,0000000000	0,0000000063	0,0004191143	0,0004191206	0,0000000000	0,0000150956	0,9999849044	1	Low	Low
44	0,0000000000	0,0000000558	0,0000191639	0,0000192197	0,0000000000	0,0029034274	0,9970965726	1	Low	Low
45	0,0000000848	0,0000000002	0,0000282232	0,0000283082	0,0029953013	0,0000066222	0,9969980765	1	Low	Low
46	0,0000000000	0,0000004185	0,0001810988	0,0001815173	0,0000000000	0,0023056888	0,9976943112	1	Low	Low
47	0,0000000000	0,0000052083	0,0001354713	0,0001406796	0,0000000000	0,0370222825	0,9629777175	1	Low	Medium
48	0,0000000000	0,0000003163	0,0000461834	0,0000464997	0,0000000000	0,0068031235	0,9931968765	1	Low	Low
49	0,0001483851	0,0005092539	0,0000000003	0,0006576393	0,2256329442	0,7743666110	0,0000004448	1	Medium	Medium
50	0,0000000000	0,0000000008	0,0000000000	0,0000000008	0,0000000000	1,0000000000	0,0000000000	1	Medium	Medium
51	0,0000000727	0,0000000000	0,0000000000	0,0000000727	1,0000000000	0,0000000000	0,0000000000	1	High	High

Gambar 3: Hasil Prediksi

d. Akurasi Prediksi

TRUE	FALSE	Accuracy
41	9	82
Pred	Dataset	
High	7	11
Medium	24	21
Low	19	18

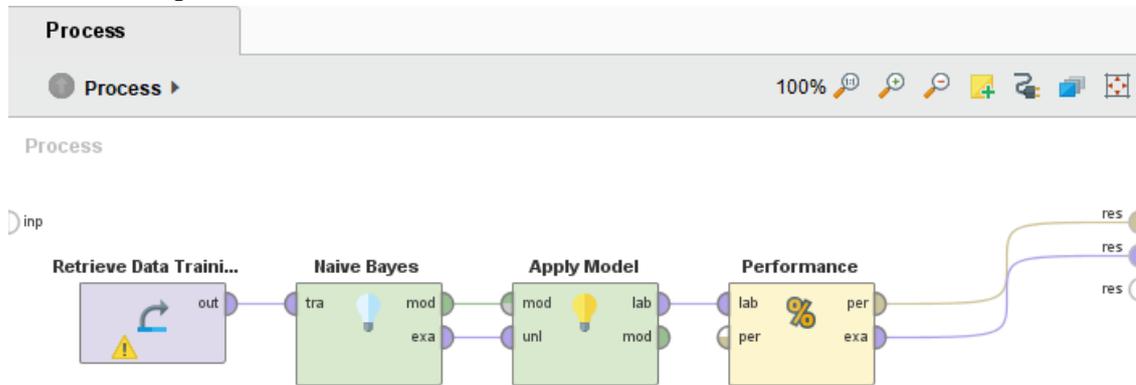
Gambar 4: Akurasi Prediksi

2. Perhitungan RapidMiner

Perhitungan yang dilakukan oleh aplikasi RapidMiner dengan

memasukkan *dataset* yang sama menggunakan algoritma Naïve Bayes yang sudah tersedia.

a. Proses RapidMiner



Gambar 5: Proses RapidMiner

b. Hasil Prediksi RapidMiner

Row No. ↑	imdb_score	prediction(L...	confidence(...	confidence(...	confidence(...	duration	director_fac...	actor_1_fac...	actor_2_fac...	actor_3_fac...
1	High	Medium	0.195	0.805	0.000	Long	Medium	High	High	High
2	High	High	0.856	0.134	0.009	Long	High	Medium	Low	Low
3	Medium	Medium	0.153	0.846	0.000	Moderate	Medium	Medium	High	Low
4	Medium	Medium	0.334	0.660	0.006	Long	High	Medium	Low	Medium
5	High	High	1	0	0	Moderate	High	High	High	Medium
6	Low	Medium	0.041	0.955	0.004	Long	High	Medium	Medium	Medium
7	Medium	Medium	0.064	0.936	0.000	Moderate	Medium	High	High	High
8	Medium	Medium	0.153	0.846	0.000	Moderate	Medium	Medium	High	Low
9	Low	Medium	0.021	0.953	0.026	Moderate	Medium	Medium	Medium	Medium
10	High	High	0.522	0.478	0.000	Short	Low	High	High	High
11	Low	Low	0	0.000	1.000	Short	Low	Low	Low	Low
12	High	Low	0.250	0.003	0.747	Moderate	Medium	Low	Low	Medium
13	High	High	0.429	0.366	0.205	Moderate	Low	High	Low	Low
14	Low	Low	0	0.014	0.986	Moderate	Low	Medium	Low	Medium
15	Medium	Medium	0	1	0	Long	High	Medium	High	High
16	Medium	Medium	0	1	0	Long	High	Medium	Medium	Medium

17	Medium	Medium	0.226	0.774	0.000	Long	High	High	High	High
18	Medium	Medium	0.195	0.805	0.000	Long	Medium	High	High	High
19	High	High	1	0	0	Long	High	High	High	High
20	High	High	1	0	0	Long	Medium	Medium	High	Medium
21	Medium	Low	0	0.433	0.567	Short	High	High	Medium	Medium
22	Medium	Medium	0.054	0.946	0.000	Moderate	High	High	Medium	Medium
23	Low	Low	0.008	0.038	0.954	Short	Low	Low	Medium	Medium
24	Low	Low	0	0.013	0.987	Short	High	Medium	Medium	Medium
25	Low	Low	0	0.268	0.732	Moderate	Low	Medium	High	High
26	Low	Low	0	0	1	Short	Low	Medium	Medium	Low
27	Medium	Medium	0	1	0	Short	Low	Low	Low	Low
28	Low	Low	0.009	0.000	0.991	Short	Low	Low	Low	Low
29	High	Low	0.366	0.000	0.633	Long	Low	Low	Low	Low
30	Low	Low	0	0.000	1.000	Short	Medium	Low	Low	Low
31	Medium	Medium	0	1	0	Moderate	Medium	High	Medium	Medium
32	Medium	Medium	0	1.000	0.000	Long	Medium	High	High	High

Gambar 6: Hasil Prediksi RapidMiner

33	Medium	Medium	0.005	0.663	0.332	Moderate	Medium	Medium	Medium	Medium
34	Low	Medium	0.006	0.993	0.001	Moderate	Medium	High	Medium	High
35	High	Medium	0.041	0.955	0.004	Long	High	Medium	Medium	Medium
36	Medium	Medium	0	0.912	0.088	Moderate	High	Medium	High	Low
37	Medium	Medium	0	0.999	0.001	Moderate	Medium	High	High	Medium
38	Medium	Medium	0	0.741	0.259	Long	High	Medium	Medium	Medium
39	Medium	Medium	0.091	0.909	0.000	Long	High	High	High	High
40	Low	Low	0	0.007	0.993	Moderate	High	Medium	Low	Low
41	Low	Low	0	0	1	Short	Low	Low	Low	Low
42	Low	Low	0	0.000	1.000	Short	Low	Medium	Low	Low
43	Low	Low	0	0.003	0.997	Short	Medium	Medium	Medium	Medium
44	Low	Low	0.003	0.000	0.997	Short	Low	Low	Low	Low
45	Low	Low	0	0.002	0.998	Short	Medium	Medium	Medium	Medium
46	Medium	Low	0	0.037	0.963	Moderate	Low	Medium	Medium	Medium
47	Low	Low	0	0.007	0.993	Moderate	High	Medium	Low	Low
48	Medium	Medium	0.226	0.774	0.000	Long	High	High	High	High
49	Medium	Medium	0	1	0	Moderate	Medium	Low	Low	Low
50	High	High	1	0	0	Moderate	Low	Low	Low	Low

Gambar 7: Hasil Prediksi RapidMiner 2

c. Akurasi Prediksi RapidMiner

Accuracy: 82.00%

	true High	true Medium	true Low	class precision
pred. High	7	0	0	100.00%
pred. Medium	2	19	3	79.17%
pred. Low	2	2	15	78.95%
class recall	63.64%	90.48%	83.33%	

Gambar 8: Akurasi Prediksi RapidMiner

3. Perhitungan Aplikasi

Perhitungan yang dilakukan pada aplikasi yang telah dibuat dengan

dataset yang sama guna mengecek apakah rumus yang digunakan sudah sesuai.

a. Hasil Prediksi Aplikasi

Tabel 6. Hasil Prediksi Aplikasi

no	High Pred	Med Pred	Low Pred	Pred. Rating	Imdb_score
1	1,112888229633 32E-04	4,583285423913 46E-04	3,65666428826036 E-10	Medium	High
2	9,496646226204 31E-05	1,488079683088 78E-05	1,04258812186879 E-06	High	High
3	5,935403891377 69E-05	3,273775302795 33E-04	1,08602929361333 E-07	Medium	Medium
4	4,748323113102 15E-05	9,374902003459 34E-05	9,12264606635195 E-07	Medium	Medium
5	6,280850678706 55E-07	0	0	High	High
6	9,496646226204 31E-06	2,187477134140 51E-04	8,10901872564618 E-07	Medium	Low
7	2,473084954740 71E-06	3,637528114217 03E-05	6,39916250445563 E-10	Medium	Medium
8	5,935403891377 69E-05	3,273775302795 33E-04	1,08602929361333 E-07	Medium	Medium
9	4,748323113102 15E-06	2,187477134140 51E-04	6,08176404423463 E-06	Medium	Low
10	8,243616515802 35E-06	7,544502755413 10E-06	2,41339843025184 E-09	High	High
11	0	6,748660694280 20E-09	6,92750935663601 E-05	Low	Low
12	7,631233574628 46E-06	1,062914059349 13E-07	2,28066151658799 E-05	Low	High
13	7,537020814447 86E-06	6,429973939272 53E-06	3,59870615929144 E-06	High	High
14	0	7,440398415443 92E-07	5,08017352819974 E-05	Low	Low
15	0	1,909702259963 94E-05	0	Medium	Medium
16	0	5,208278890810 75E-06	0	Medium	Medium
17	1,483850972844 42E-04	5,092539359903 84E-04	2,92533143060829 E-10	Medium	Medium
18	1,112888229633 32E-04	4,583285423913 46E-04	3,65666428826036 E-10	Medium	Medium
19	1,059893552031 73E-05	0	0	High	High
20	1,271872262438 08E-05	0	0	High	High
21	0	3,936718738330 12E-08	5,16199108692754 E-08	Low	Medium
22	8,793190950189	1,543193745425	1,63818560114064	Medium	Medium

	17E-06	41E-04	E-08		
23	2,261106244334 36E-07	1,080235621797 78E-06	2,73679381990559 E-05	Low	Low
24	0	2,170116204504 48E-06	1,65576026104288 E-04	Low	Low
25	0	2,652364249949 92E-06	7,25739075457106 E-06	Low	Low
26	9,793416420773 19E-07	4,219670397647 60E-07	4,19114316076479 E-03	Low	Low
27	0	1,093532982869 48E-09	0	Medium	Medium
28	1,271872262438 08E-06	1,093532982869 48E-09	1,45147815091421 E-04	Low	Low
29	7,631233574628 46E-06	4,920898422912 65E-09	1,31952559174019 E-05	Low	High
30	0	2,108956466962 56E-09	3,62869537728553 E-05	Low	Low
31	0	2,192959532972 94E-05	0	Medium	Medium
32	0	2,104569837511 28E-05	1,79976445437815 E-09	Medium	Medium
33	3,391659366501 54E-07	4,557244029459 40E-05	2,28066151658799 E-05	Medium	Medium
34	2,119787104063 46E-07	3,472185927207 16E-05	3,83949750267338 E-08	Medium	Low
35	9,496646226204 31E-06	2,187477134140 51E-04	8,10901872564618 E-07	Medium	High
36	0	4,546910142771 29E-05	4,39841863913398 E-06	Medium	Medium
37	0	5,846027326420 22E-05	6,47915203576133 E-08	Medium	Medium
38	0	1,953104584054 03E-05	6,84198454976396 E-06	Medium	Medium
39	1,059893552031 73E-05	1,060945699979 97E-04	1,09699928647811 E-09	Medium	Medium
40	0	3,163434700443 85E-07	4,61833957109068 E-05	Low	Low
41	0	0	9,48678530009289 E-07	Low	Low
42	0	6,326869400887 69E-09	4,19114316076479 E-04	Low	Low
43	0	5,580298811582 94E-08	1,91638919102185 E-05	Low	Low
44	8,479148416253 85E-08	1,874627970633 39E-10	2,82231862677764 E-05	Low	Low
45	0	4,185224108687 21E-07	1,81098778551565 E-04	Low	Low

46	0	5,208278890810 74E-06	1,35471294085327 E-04	Low	Medium
47	0	3,163434700443 85E-07	4,61833957109068 E-05	Low	Low
48	1,483850972844 42E-04	5,092539359903 84E-04	2,92533143060829 E-10	Medium	Medium
49	0	7,991835032700 24E-10	0	Medium	Medium
50	7,267841499646 15E-08	0	0	High	High
Prediksi Benar:					41
Prediksi Salah:					9
Akurasi					82.00%

Dengan menggunakan variabel yang sesuai dan dilakukannya proses *data cleansing* diperoleh akurasi yang lebih baik yaitu sebesar 53,99%. Untuk mendapatkan hasil terbaik dalam prediksi tidak dapat terpaku pada satu algoritma saja, setiap *dataset* memiliki keunikan dan kebutuhannya sendiri sehingga diperlukannya melakukan perbandingan beberapa algoritma *data mining* terhadap sebuah *dataset* yang akan digunakan. Terbukti pada penelitian ini bahwa algoritma Naïve Bayes lebih cocok dengan *dataset* yang digunakan.

IV. SIMPULAN

Algoritma yang digunakan dalam aplikasi prediksi adalah Naïve Bayes dengan tingkat keakuratan sebesar 53,99% menggunakan *dataset* “imdb 5000 movie dataset” yang dibuat oleh Yueming dari Northeastern University. Dengan aplikasi yang telah dibuat *User* dapat mengetahui prediksi *rating* film guna membantu dalam proses produksi dan distribusi film *user*.

REFERENSI

- [1] Pratiwi, Riszky Wijayatun dan Nugroho, Yusuf Sulisty, “Prediksi Rating Film Menggunakan Metode Naïve Bayes”, Jurnal Teknik Elektro, vol. 8 no. 2, 2016, hh. 61.
- [2] Oxford dictionaries. English Oxford Living Dictionaries, 2018. [Online]. Tersedia: <https://en.oxforddictionaries.com/definition/rating> [Diakses: Des. 17, 2018].
- [3] Herdianto, “Prediksi Kerusakan Motor Induksi Menggunakan Jaringan Saraf Tiruan Backpropagation”, S2 tesis, Fakultas Teknik, Universitas Sumatera Utara, Medan, Indonesia, 2013.
- [4] Yustanti, Wiyli, “Algoritma *K-Nearest Neighbour* untuk Memprediksi Harga Jual Tanah”, Jurnal Matematika, Statistika & Komputasi, vol. 9, no.1, 2012, 57-68.
- [5] Silalahi, Mesri, “Analisis Clustering Menggunakan Algoritma K-Means Terhadap Penjualan Produk Pada PT. Batamas Niaga Jaya”, Jurnal CBIS, vol. 06, no 02, 2018, 20-35.

BIOGRAFI

Andrew Ryan Yosafat Lulus pada Program Studi Teknik Informatika (S1) Universitas Buddhi Dharma di bidang Basis Data, 2019.

Yusuf Kurnia Lulus pada Program Studi Sistem Informasi (S1) STMIK Buddhi Dharma di bidang Perusahaan, 2009, Program Studi Ilmu Komputer (S2) Universitas Budi Luhur di bidang Teknologi Sistem Informasi, 2013. Saat ini sebagai dosen tetap di Universitas Buddhi Dharma.