



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB II

LANDASAN TEORI

2.1 Film

Film menurut Kamus Besar Bahasa Indonesia memiliki arti yaitu lakon atau cerita gambar hidup, dalam Bahasa Inggris film biasanya disebut sebagai *movie*. Film merupakan kombinasi dari kumpulan gambar diam yang ketika ditampilkan pada sebuah layar akan menciptakan ilusi menjadi gambar yang bergerak [6]. Film juga menjadi sarana baru yang dipakai untuk menyebarkan suatu hiburan yang telah menjadi kebiasaan sejak lama serta menampilkan cerita peristiwa, drama, musik, dan lainnya pada masyarakat umum [5].

Film dapat dibedakan jenis-jenisnya. Secara luas film dapat dikategorikan sebagai: fiksi yaitu tidak nyata atau dibuat-buat, non-fiksi yaitu berdasarkan kisah atau kejadian nyata, juga ada yang merupakan kegabungan dari keduanya. Jika melihat dari *genre*, film memiliki berbagai jenis *genre* yaitu:

- a. Aksi adalah *genre* film dengan cerita yang menegangkan dengan melibatkan adegan seperti tembak-tembakan, pertarungan, dan sebagainya. *Genre* ini biasanya akan menceritakan tentang kebaikan melawan kejahatan.
- b. Petualangan adalah *genre* film yang menceritakan tokoh utama yang melakukan perjalanan untuk mendapatkan sesuatu atau mencapai sebuah tujuan.
- c. Animasi adalah *genre* film yang menggunakan karakter kartun atau bukan orang sebenarnya sebagai tokoh maupun latarnya.
- d. Komedi adalah *genre* film yang menceritakan hal-hal lucu dan biasanya akan membuat penonton tertawa.
- e. Dokumenter merupakan film yang menceritakan kisah hidup seseorang atau juga kejadian nyata yang terjadi di masa lalu.
- f. Drama adalah *genre* film dimana biasanya akan menceritakan kisah-kisah yang melibatkan emosi untuk menontonnya.

- g. Horror adalah *genre* film yang menakutkan. Film horror biasanya akan melibatkan hantu, suasana yang gelap dan mencengangkan untuk membuat suasana yang menyeramkan.
- h. Romantis adalah *genre* dimana film akan menceritakan kisah percintaan dari tokoh-tokoh didalamnya.

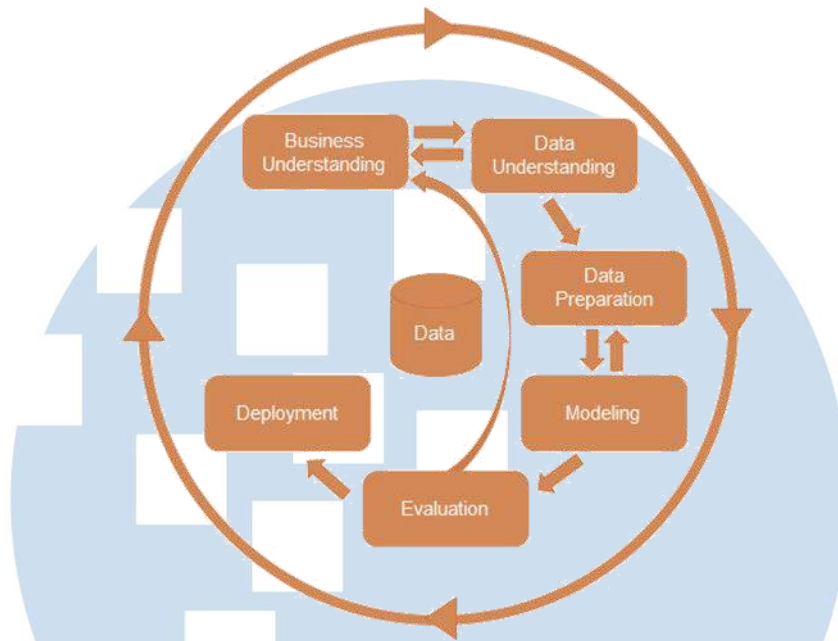
2.2 **Rating Film**

Rating adalah elemen penting untuk menentukan kualitas dari sebuah film. *Rating* merupakan kesimpulan yang menggambarkan kualitas dari elemen-elemen yang ada dalam sebuah film [4]. *Rating* dari sebuah film bisa didapatkan berdasarkan *feedback* dari masyarakat yang telah menonton film tersebut [7]. *Rating* yang tinggi tentunya akan lebih menarik masyarakat lainnya untuk menonton film tersebut dan sebaliknya untuk film dengan *rating* yang rendah akan lebih jarang ditonton oleh masyarakat.

2.3 **CRISP-DM**

Cross Industry Standard Process for Data mining atau yang biasanya dikenal dengan singkatan CRISP-DM adalah salah satu metode yang populer untuk meningkatkan kesuksesan model *data mining*. Metode CRISP-DM berbentuk siklus yang terdiri atas enam tahap yang akan membantu implementasi dari *data mining* untuk diaplikasikan dalam lingkungan yang nyata seperti membantu untuk pengambilan keputusan bisnis [11]. Enam tahap yang ada dalam siklus metode CRISP-DM menurut [12] yaitu:

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



Gambar 2. 1 CRISP-DM [12]

- a. *Business Understanding* merupakan proses tahapan pertama dari metode CRISP-DM. Pada tahap ini akan dilakukan pemahaman bisnis dari proyek yang dikerjakan dengan melakukan penentuan tujuan dan kebutuhan proyek, serta menyiapkan strategi awal untuk mencapai tujuan yang diinginkan.
- b. *Data Understanding* merupakan proses tahapan kedua dari metode CRISP-DM. Pada tahap ini akan dilakukan pemahaman dari data yang digunakan. Proses ini mencakup pengumpulan data, menganalisis data, dan mengevaluasi kualitas data.
- c. *Data Preparation* merupakan proses tahapan ketiga dari metode CRISP-DM. Pada tahap ini akan dilakukan pengolahan data yang mencakup menyiapkan data yang tersedia, memilih kasus dan variable yang akan dianalisis, perubahan pada variabel, dan menyiapkan data awal yang siap untuk dilakukan permodelan.
- d. *Data Modeling* adalah tahap keempat dari metode CRISP-DM. Pada tahap ini akan dilakukan pembuatan model dan diaplikasikan terhadap data yang sudah disiapkan. Model juga akan dioptimalkan dan jika diperlukan, proses dapat kembali ke tahap pengolahan data

untuk menjadikan data kedalam bentuk yang sesuai dengan teknik yang akan digunakan.

- e. *Evaluation* adalah tahap kelima dari metode CRISP-DM. Tahap ini dilakukan untuk mengevaluasi model yang digunakan dalam tahap modeling untuk mendapatkan kualitas dan efektivitas sebelum disebarakan untuk digunakan. Evaluasi juga dilakukan dengan tujuan menyesuaikan model yang dibuat dengan tujuan yang ada dalam tahap Business Understanding.
- f. *Deployment* adalah tahap keenam dari metode CRISP-DM. Pada tahap ini akan dilakukan penyusunan laporan atau presentasi dari informasi yang didapat dari evaluasi pada proses *data mining*.

2.4 Naïve Bayes

Naïve Bayes merupakan Teknik pengelompokan statistik yang bisa digunakan untuk memprediksi probabilitas anggota dari suatu *class* [5]. Naïve Bayes juga memiliki akurasi dan tingkat kecepatan yang sangat kuat ketika diaplikasikan pada *database* dengan *big data* menurut [7]. Berikut merupakan rumus dari Naïve Bayes:

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

Rumus 2. 1 Naïve Bayes

Keterangan rumus 2.1 :

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

2.5 KNN (K-Nearest Neighbor)

Algoritma KNN atau K-Nearest Neighbor merupakan salah satu metode untuk melakukan klasifikasi terhadap objek berdasarkan pada data pembelajaran dengan jarak paling dekat dengan objek tersebut. Data pembelajaran akan diproyeksikan ke ruang berdimensi banyak dimana masing-masing dimensi akan merepresentasikan fitur dari data [7].

Berikut merupakan rumus dari algoritma KNN:

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

Rumus 2. 2 KNN

Keterangan rumus 2.2:

$x_{training}^i$: Data training ke-i

$x_{testing}$: Data testing

I : Baris ke- dari tabel

n : Jumlah data training

2.6 Feature selection

Feature selection merupakan teknik *preprocessing* dalam *data mining* untuk memilih fitur yang relevan dan mengurangi data dengan menghapus atribut yang tidak penting dan redundan dari dataset [10]. Tujuan dari *feature selection* adalah untuk mendapatkan performa prediksi yang lebih baik, mempercepat proses prediksi, mengurangi biaya, dan mengetahui proses dari menghasilkan data yang lebih baik [10].

2.7 Forward selection

Forward selection merupakan salah satu metode dari teknik *feature selection*. *Forward selection* bertujuan untuk menambah variabel yang

dikendalikan secara 1 per 1 ke dalam persamaan [13]. *Forward selection* dimulai dari himpunan kosong lalu menambahkan variabel 1 per 1 dan dievaluasi hingga semua variabel telah diujikan [13].

2.8 *Backward elimination*

Backward elimination juga merupakan salah satu metode dari teknik *feature selection*. *Backward elimination* bekerja dengan cara kebalikan dari metode *forward selection* [13]. Proses *backward elimination* dimulai dengan memasukan semua variabel yang ada dan mengeliminasi 1 per 1 variabel yang dianggap tidak signifikan berdasarkan nilai akurasi atau error, proses berlanjut hingga tidak ada lagi variabel yang tidak signifikan [13].

2.9 **RapidMiner**

Tools yang dipakai dalam pengerjaan penelitian ini adalah *software* RapidMiner. RapidMiner dapat mengimpor informasi dari berbagai sumber database yang kemudian dianalisa dalam sebuah aplikasi [5]. RapidMiner merupakan salah satu *tool data mining* umum yang dikembangkan oleh perusahaan RapidMiner di Jerman. RapidMiner menjadi populer di beberapa tahun belakangan ini dan memiliki komunitas yang besar [14].

RapidMiner menawarkan lingkungan yang terintegrasi dengan GUI yang menarik secara visual dan ramah pengguna. RapidMiner difokuskan pada proses yang mungkin berisi subproses. Proses mengandung operator dalam bentuk komponen visual. Operator adalah implementasi dari algoritma DM dan sumber data.. Aliran data dibangun oleh *drag-and-drop* operator dan dengan menghubungkan input dan output dari operator yang sesuai. RapidMiner juga menawarkan opsi *wizard* aplikasi yang membangun proses secara otomatis berdasarkan tujuan proyek yang diperlukan (misalnya pemasaran langsung, analisis churn, analisis sentimen). Ada juga tutorial yang tersedia untuk banyak tugas tertentu sehingga alat ini memiliki kurva belajar yang stabil [14].

2.10 Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

Judul Jurnal	Penulis / Tahun	Nama Jurnal	Hasil	Simpulan
Penerapan <i>Feature Selection</i> untuk Prediksi Lama Studi Mahasiswa	I. Adnyana / 2019 [15]	JURNAL SISTEM DAN INFORMATIKA Vol. 13, No. 2, Mei 2019.	Meningkatkan akurasi Naïve Bayes dari 69,95% menjadi 77,83% menggunakan <i>feature selection</i> .	Penerapan <i>feature selection</i> secara umum dapat meningkatkan akurasi dari algoritma klasifikasi Naïve Bayes.
Aplikasi Prediksi <i>Rating Film</i> dengan Perbandingan Metode <i>Naïve Bayes</i> dan KNN Berbasis Website Menggunakan <i>Framework Codeigniter</i>	A. Yosafat, Y. Kurnia / 2019 [7]	JURNAL ALGOR - VOL. 1 NO. 1 (2019)	Naïve Bayes 53,99% KNN 51,26%	Algoritma yang digunakan dalam aplikasi adalah algoritma Naïve Bayes karena memiliki tingkat akurasi yang lebih tinggi yaitu 53,99% dibanding KNN 51,26%
<i>A Review of Feature Selection and Classification Approaches for Heart Disease Prediction</i>	F. Firdaus, H. Nugroho, I. Susanti / 2021 [10]	IJITEE, Vol. 4, No. 3, September 2020	<i>Feature Selection</i> dapat digunakan untuk memilih fitur yang paling relevan dan meningkatkan performance dari model yang dibuat.	<i>Feature Selection</i> dapat digunakan untuk penelitian ini sehingga dapat meningkatkan akurasi model yang dibuat.

Judul Jurnal	Penulis / Tahun	Nama Jurnal	Hasil	Simpulan
Implementasi Model Crisp-Dm Untuk Menentukan <i>Sales Pipeline</i> Pada Pt X	D. Feblian, D. Daihani / 2017 [12]	JURNAL TEKNIK INDUSTRI, Vol 6, No. 1, 2017	CRISP-DM dapat menciptakan lingkungan bisnis yang efisien dalam mendapatkan informasi yang dibutuhkan.	Berdasarkan hasil yang didapatkan dalam penelitian terdahulu, CRISP-DM cocok untuk digunakan dalam penelitian ini.

Berdasarkan penelitian terdahulu pada tabel 2.1, mendapatkan hasil bahwa algoritma Naïve Bayes dan KNN merupakan algoritma yang populer dalam melakukan prediksi rating dan teknik *Feature Selection* untuk meningkatkan akurasi model serta CRISP-DM sebagai metode pengerjaan yang cocok. Namun hasil yang didapatkan dalam penelitian-penelitian tersebut masih terbilang kurang baik sehingga dijadikan dasar dari peningkatan kinerja algoritma yang akan dilakukan dalam penelitian ini.

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA