

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Penelitian Terdahulu**

Penelitian sebelumnya yang dilakukan oleh Bernadus Gunawan Sudarsono, Marcell Ignatius Leo, dan Felix Hendrawan dengan judul “ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPIDMINER” adalah bertujuan memperkirakan genre film yang paling banyak ditonton/ditayangkan oleh pengguna Netflix menggunakan aplikasi Rapidminer.

Persamaan dari penelitian yang dilakukan oleh Bernadus Gunawan Sudarsono, Marcell Ignatius Leo, Ali santoso, dan Felix Hendrawan adalah sama-sama melakukan proses analisis data mining terhadap pengguna netflix.

Perbedaan dari penelitian terdahulu adalah dari segi metode dan pengumpulan data. Jika penelitian terdahulu menggunakan metode K-fold Cross Validation yaitu metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi / evaluasi. Mengumpulkan data pengguna netflix melalui internet. Selain itu, pengujian data pengguna netflix menggunakan aplikasi Rapidminer , tetapi penelitian ini menggunakan metode K-means Clustering yaitu salah satu metode *cluster analysis* non hirarki yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya.

Pengumpulan data pengguna netflix melalui kuesioner dengan objek penelitian mahasiswa program studi DKV Universitas Pasundan Bandung, karena mahasiswa tersebut belum pernah dilakukan penelitian yang serupa.

## **2.2 Definisi Minat**

Menurut Sardiman minat diartikan sebagai suatu kondisi yang terjadi apabila seseorang melihat ciri-ciri atau arti sementara situasi yang dihubungkan dengan keinginan-keinginan atau kebutuhan-kebutuhan sendiri. Dari pengertian Sardiman tersebut dapat kita peroleh bahwa minat itu adalah ciri- ciri atau arti sementara situasi yang dihubungkan dengan keinginan atau kebutuhan. Dimana dapat kita simpulkan bahwa minat adalah keinginan atau kebutuhan.

Menurut Kamus Besar Bahasa Indonesia (KBBI), minat adalah kecenderungan hati yang tinggi terhadap sesuatu, gairah, dan keinginan. Sehingga dari pendapat KBBI tersebut dapat kita simpulkan bahwa minat memiliki 3 pengertian, yaitu: kecenderungan, gairah dan keinginan. Sehingga minat adalah kecenderungan, gairah dan keinginan. Minat adalah keinginan sama seperti pengertian menurut Sardiman. [4]

## **2.3 Definisi Analisis**

Analisis didefinisikan sebagai metode penelitian yang objektif, sistematis, dan bersifat kuantitatif terhadap isi pesan yang *manifest* . Berdasarkan definisi ini jelas bahwa analisis isi memiliki karakteristik utama yakni objektif, sistematis, dan bersifat kuantitatif. Analisis isi juga hanya mencermati hal yang bersifat manifest atau terlihat. Metode ini tidak membahas ideologi, wacana, atau makna dari teks yang diteliti.[5]

Analisis memiliki karakteristik objektif Hal ini bermakna bahwa penelitian analisis isi berusaha menyajikan cerminan dari pesan yang disampaikan tanpa memberikan pretensi atau sudut pandang tertentu dari peneliti. Penilaian terhadap hasil penelitian merupakan otoritas dari pembaca karena analisis isi tidak berusaha memberikan bingkai pemaknaan terhadap data yang diperoleh. Peneliti juga hanya bertugas untuk menyampaikan isi. Ia tidak diperkenankan untuk menghadirkan subjektifitas yang dimiliki.[5]

#### **2.4 Definisi Data Mining**

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. [6]

Pengertian data mining dapat didefinisikan dengan cara pandang dan pengetahuan yang berbeda sehingga para ahli memberikan beberapa definisi tentang data mining. Beberapa definisi yang terkenal tentang pengertian data mining adalah sebagai berikut:

a) Data mining adalah serangkaian proses menggali atau menambang suatu kumpulan data untuk menghasilkan pengetahuan yang selama ini tidak diketahui secara manual.

b) Data mining atau *Knowledge Discovery in Database* (KDD) merupakan proses pengambilan informasi yang tersembunyi, di mana informasi tersebut sebelumnya tidak dikenal. Proses dalam KDD meliputi proses pendekatan secara

teknis seperti clustering, klasifikasi dengan meliputi metode yang merupakan irisan *artificial intelligence (AI)*, *machine learning (ML)* dan statistik. dapat menangani data berskala besar serta data mining dapat memanfaatkan data pengalaman masa lampau untuk meningkatkan proses model pembelajarannya seperti pada penerapan klasifikasi.[7]

#### **2.4.1 Tahapan Data mining**

Adapun penjelasan mengenai proses tahapan dalam *Knowledges Discovery in Database* yang akan di jelaskan sebagai berikut.

##### a) Data Selection

Seleksi atau pemilihan data perlu dilakukan dari himpunan data operasional sebelum tahapan penggalian informasi dalam *Knowledge Discovery in Databases* dimulai. Hasil data yang sudah diseleksi disimpan dalam suatu berkas dan akan dipergunakan dalam proses data mining, di mana letaknya terpisah dari basis data operasional.

Selain itu, tahapan ini menciptakan himpunan data target, pemilihan himpunan data dan fokus pada sub set data sampel atau variabel di mana proses *discovery* akan dilaksanakan.

##### b) Pre-processing (*Cleaning*)

Tahapan pembersihan data ini merupakan operasi dasar. Sebelum proses data mining dilaksanakan, perlu dilakukan proses *cleaning* pada data terfokus dari KDD tersebut. Proses dari *cleaning* ini meliputi beberapa aktivitas, seperti membuang duplikasi data, memeriksa data yang dinilai tidak konsisten serta mencari dan memperbaiki kesalahan yang ada pada data seperti kesalahan cetak

(*typography*). Selain itu juga perlu dilaksanakan *enrichment* atau proses “memperkaya” data sebelumnya dengan data yang lebih relevan serta diperlukan dalam KDD, contohnya data atau informasi yang bersifat eksternal.

c) *Transformation*

Tahapan *transformation* yaitu proses pencarian berbagai fitur yang dapat digunakan untuk mempresentasikan data, tergantung pada sasaran (*goal*) yang ingin dituju. Tahapan *coding* ini juga merupakan proses transformasi terhadap data yang sudah diseleksi, hingga data tersebut dapat sesuai digunakan dalam proses data mining, dengan cara memodifikasi data ke dalam model analitis dan membuat model sesuai analisa yang diharapkan dan diperlukan dalam algoritma data mining. Proses ini bergantung pada pola dan jenis informasi yang dicari dalam *database* (basis data).

d) Data Mining

Data mining di sini dapat diartikan sebagai proses menemukan pola dan informasi yang unik di dalam data yang telah diseleksi sebelumnya dengan menggunakan metode atau teknik tertentu. Adapun metode atau algoritma yang digunakan dalam data mining sendiri sangat variatif, di mana pemilihan teknik atau algoritma yang tepat sangat bergantung pada keseluruhan proses serta tujuan dari KDD.

e) *Interpretation / Evaluation*

Tahapan ini adalah suatu bagian dari proses *Knowledge Discovery in Databases* meliputi pemeriksaan terhadap pola ataupun informasi yang ditemukan, apakah bertentangan dengan hipotesa atau fakta yang telah ada sebelumnya. Pada

tahapan *interpretation / evaluation* ini, juga dilakukan proses terjemah dari beberapa pola yang didapatkan dari data mining. Pola informasi yang dihasilkan dari proses ini harus ditampilkan ke dalam bentuk yang lebih mudah dipahami oleh berbagai pihak yang berkepentingan.

f) Knowledge

Tahap terakhir ini merupakan presentasi dari pola yang telah ditemukan untuk mendapatkan aksi pada tahap terakhir dari proses data mining ini adalah bagaimana cara memformulasikan aksi atau keputusan dari hasil analisis yang sudah didapatkan. Hasil mining yang telah melalui tahap evaluasi digunakan dalam sistem informasi perusahaan dan perilaku organisasi.[7]

## 2.5 Definisi Film

Film dalam kamus besar Bahasa Indonesia, memiliki arti sebagai selaput tipis yang terbuat dari seluloid yang berfungsi sebagai tempat gambar negatif (yang akan dibuat potret) maupun gambar positif (yang akan dimainkan di bioskop). Selain itu, film juga diartikan sebagai lakon (cerita) gambar hidup. Sebagai bagian dari industri, film juga memiliki arti sebagai sesuatu bagian dari produksi ekonomi di suatu masyarakat dan film mesti dipandang dalam hubungannya dengan produk-produk lainnya. Film juga termasuk bagian dari komunikasi yang merupakan bagian terpenting dari sebuah sistem yang digunakan oleh individu maupun kelompok yang berfungsi untuk mengirim dan menerima pesan.[8]

Secara harfiah film adalah *cinematographie*. *Cinematographie* berasal dari kata *cinema* yang memiliki arti "gerak". *Tho* atau *phytos* yang memiliki arti (cahaya). Oleh karena itu, film juga dapat diartikan sebagai melukis sebuah gerak

dengan memanfaatkan cahaya. Selanjutnya, Menurut Ibrahim film juga memiliki arti sebagai dokumen sosial dan budaya yang membantu mengkomunikasikan zaman ketika film itu dibuat bahkan sekalipun ia tak pernah dimaksudkan untuk itu. Menurut Javadalasta juga menyatakan bahwa film merupakan rangkaian dari gambar yang bergerak dan membentuk suatu cerita yang dikenal dengan sebutan *movie* atau video. Film sebagai media *audio visual* yang terdiri dari potongan gambar yang disatukan menjadi kesatuan utuh, dan memiliki kemampuan dalam menangkap realita sosial budaya, tentu membuat film mampu menyampaikan pesan yang terkandung di dalamnya dalam bentuk media *visual*. Pada film "Lamaran" terlihat dengan jelas bahwa sutradara mengonstruksi sebuah realita sosial budaya yang ada di Indonesia. Realita sosial budaya tersebut dikemas dengan sebaik mungkin agar film "Lamaran" mampu berkomunikasi dengan penonton tanpa ada batas.[8]

## **2.6 Definisi K-means Clustering**

Algoritma K-Means klustering merupakan metode non-hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat kluster awal. Pada tahap ini pusat kluster dipilih secara acak dari sekumpulan populasi data. Berikutnya K-Means menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat kluster yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap kluster. Posisi pusat kluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap pusat kluster dan terakhir akan terbentuk posisi pusat kluster yang baru.

Algoritma klastering K-Means dapat membagi data berdasarkan jarak antar data pada kelompok yang telah ditetapkan. Algoritma ini bergantung pada fungsi untuk mengukur data yang mempunyai ciri khas sama. Jarak itu sendiri dihitung menggunakan fungsi euclidean. Kemudian data dimasukkan dalam kelompok yang mempunyai jarak terdekat.[9]

Langkah-langkah pengelompokan data adalah :

1. Pilih jumlah klaster.
2. Inisialisasi awal dan pusat klaster dilakukan secara random.
3. Setiap data ditempatkan ke pusat klaster terdekat berdasarkan jarak antar obyek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan Metode Jarak Euclidean.

(*Euclidean Distance*) dengan rumus seperti di bawah ini:

$$d_{ij} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \dots\dots\dots(1)$$

**Gambar 2. 1 Formula Euclidean**

Dimana :

$X_{ij}$  = Pusat Cluster

$C_{kj}$  = Data

4. Penentuan pusat cluster baru

Setelah mengelompokkan data ke dalam pusat duster awal yang terdekat, maka langkah selanjutnya menentukan pusat cluster baru dengan cara dihitung berdasarkan data anggota tiap-tiap cluster yang

sudah didapatkan dengan menggunakan rumus sebagai berikut :

$$C_i = \frac{1}{m} \sum_{j=1}^m X_j \quad \dots\dots\dots(2)$$

**Gambar 2. 2 Rumus Penentuan Cluster baru**

Dimana : dilakukan penjumlahan data yang menjadi anggota clusternya dibagi sebanyak data anggota clusternya, maka akan didapatkan titik pusat duster yang baru.

5. Hitung kembali jarak tiap objek dengan pusat klaster yang baru, hingga klaster tidak berubah, maka proses pengklasteran selesai.[9][17]

## **2.7 Definisi Aplikasi**

Aplikasi yaitu perangkat lunak yang digunakan untuk membantu pemakai komputer untuk melaksanakan pekerjaannya. Jika ingin mengembangkan program aplikasi sendiri, maka untuk menulis program aplikasi tersebut, dibutuhkan suatu bahasa pemrograman, yaitu language software, yang dapat berbentuk *assembler*, *compiler* ataupun *interpreter*. Sekarang, banyak sekali program-program aplikasi yang tersedia dalam bentuk paket-paket program. Ini adalah program-program aplikasi yang sudah ditulis oleh orang lain atau perusahaan-perusahaan perangkat lunak. Beberapa perusahaan perangkat lunak telah memproduksi paket-paket perangkat lunak yang mempunyai reputasi internasional.[10]

## **2.8 Definisi Kuesioner**

Kuesioner atau daftar isian adalah satu set pertanyaan yang tersusun secara

sistematis dan standar sehingga pertanyaan yang sama dapat diajukan terhadap setiap responden. Sistematis yang dimaksud adalah bahwa item-item pertanyaan disusun menurut logika (*logical sequence*) sesuai dengan maksud dan tujuan pengumpulan data. Sedangkan yang dimaksud standar adalah setiap item pertanyaan mempunyai pengertian, konsep, dan definisi yang sama.

Usaha untuk membuat kuesioner suatu survei yang baik, harus diarahkan pada dua tujuan utama, yaitu:

- Memperoleh informasi/data yang berhubungan dengan maksud dan tujuan survei.
- Mengumpulkan informasi dengan kecermatan dan ketelitian yang dapat dipertanggungjawabkan.

Agar memenuhi tujuan pertama, rancangan kuesioner harus benar-benar sesuai dengan situasi di mana lingkup topik yang diselidiki dapat dibatasi. Informasi yang dikumpulkan harus berupa fakta dan bersifat objektif sesuai dengan tujuan survei. Selain itu, pertanyaan-pertanyaan tersebut hanya ditujukan kepada responden yang diketahui berhak dan mampu menjawabnya.[11]

## **2.9 Populasi dan Sampel**

### **2.9.1 Populasi**

Populasi adalah kumpulan dari keseluruhan pengukuran, objek. atau individu yang sedang dikaji. Jadi pengertian populasi dalam statistik tidak terbatas pada sekelompok/kumpulan orang-orang, namun mengacu pada seluruh ukuran, hitungan, atau kualitas yang menjadi fokus perhatian suatu kajian. Suatu pengamatan/survey terhadap seluruh anggota populasi disebut sensus.[12] Pada

penelitian ini yang menjadi populasinya adalah mahasiswa program studi DKV Fakultas Ilmu Seni dan Sastra Universitas Pasundan Bandung yang berlangganan di aplikasi Netflix.

### **2.9.2 Sampel**

Sampel adalah sebagian atau wakil populasi yang diteliti. Pendapat yang senada pun dikemukakan oleh Sugiyono. Ia menyatakan bahwa sampel adalah sebagian dari jumlah dan karakteristik yang dimiliki oleh populasi. Bila populasi besar, dan peneliti tidak mungkin mempelajari semua yang ada pada populasi, Misalnya karena keterbatasan dana, tenaga dan waktu, maka peneliti dapat menggunakan sampel yang diambil dari populasi itu. Apa yang dipelajari dari sampel itu, kesimpulannya akan diberlakukan untuk populasi. Untuk itu sampel yang diambil dari populasi harus betul-betul representative.[12]

Margono menyatakan bahwa sampel adalah sebagai bagian dari populasi, sebagai contoh yang diambil dengan menggunakan cara-cara tertentu. Hadi menyatakan bahwa sampel dalam suatu penelitian timbul disebabkan hal berikut:

- Peneliti bermaksud mereduksi objek penelitian sebagai akibat dari besarnya jumlah populasi, sehingga hanya meneliti sebagian saja.
- Penelitian bermaksud mengadakan generalisasi dari hasil-hasil penelitiannya, dalam arti mengenakan kesimpulan-kesimpulan kepada objek, gejala atau kejadian yang lebih luas.[14]