

ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER

ANALYSIS DATA MINING NETFLIX DATA USING THE RAPID MINER APPLICATION

Bernadus Gunawan Sudarsono^{1)*}, Marcell Ignatius Leo¹⁾, Ali santoso³⁾ dan Felix Hendrawan⁴⁾

¹⁾ Program Studi Sistem Informasi, Universitas Bung Karno, Jakarta
^{2,3,4,5)} Program Studi Sistem Informasi, Universitas Bunda Mulia, Jakarta

Diterima 02 Maret 2021 / Disetujui 10 Maret 2021

ABSTRACT

Netflix is a streaming platform that presents Netflix original films and series, Content produced by Netflix and in collaboration with global filmmakers (Global Film Maker) as an advantage compared to other streaming subscriptions. Netflix invites viewers to get to know global culture with its original films. In Indonesia, Netflix officially and premierly released the original Netflix movie with the title "The Night Comes for Us". This study aims to show how the footage of the film "The Night Comes for Us" displays and reflects the image of Indonesia based on Jean Baudrillard's imaging model that starts from the stages of representation, ideology, simulation and simulakra. Roland Barthes's semiotic analysis is used in this study as a bridge to understand the first and second phases of Baudrillard's imaging. The findings obtained explain that the image of Indonesia displayed in the excerpt of the research object does not represent any reality, but is a mere formation. The image displayed is nothing but a commodity used to seduce the audience. Imaging simulations are based on the motivation to "win" certain market segments.

Keywords: Netflix, Platform Streaming, Data Mining

ABSTRAK

Netflix adalah sebuah platform streaming yang menyajikan Netflix original movie dan series, Konten yang diproduksi oleh Netflix dan bekerjasama dengan sineas global (*Global Film Maker*) sebagai keunggulan dibandingkan dengan langganan streaming lainnya. Netflix mengajak penontonnya untuk mengenal kebudayaan global dengan film-film originalnya. Di Indonesia, Netflix secara resmi dan perdana ini merilis Netflix original movie dengan judul "The Night Comes for Us". Penelitian ini bertujuan untuk menunjukkan bagaimana cuplikan film "The Night Comes for Us" menampilkan dan merefleksikan citra Indonesia dengan berpijak pada model pencitraan Jean Baudrillard yang dimulai dari tahapan representasi, ideologi, simulasi dan simulakra. Analisa semiotika Roland Barthes digunakan dalam penelitian ini sebagai jembatan untuk memahami fase pertama dan kedua pencitraan Baudrillard. Temuan yang didapatkan menjelaskan bahwa citra Indonesia yang di tampilan dalam cuplikan objek penelitian tidak mewakili realitas apapun, melainkan sebuah bentukan belaka. Citra yang ditampilkan tak lain adalah sebuah komoditas yang digunakan untuk merayu penonton. Simulasi pencitraan didasari dengan motivasi untuk "memenangkan" segmentasi pasar tertentu.

Kata Kunci: Netflix, Platform Streaming, Data Mining.

*Korespondensi Penulis:

E-mail: gunawanbernadus@ubk.ac.id

PENDAHULUAN

Netflix adalah layanan online berbasis langganan yang memungkinkan para member yang ada menonton acara TV dan film tanpa iklan di perangkat yang terhubung ke Internet atau dengan kata lain adalah layanan yang memungkinkan pengguna atau penonton, menonton tayangan kesukaan mereka di mana pun, kapan pun, bisa menggunakan smartphone, smartTV, tablet, PC, dan laptop (Kompas, 2021).

Rapid miner adalah (Şeker dan Şadi, 2013) platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambangan teks, dan analisis prediktif.

Rapid miner harus mendapatkan *raw data*/ data mentah dari netflix untuk di proses lebih lanjut dan menghasilkan dataset yang baru dan lebih bersih. Jumlah data yang begitu besar yang terdapat dalam database netflix mengakibatkan sulitnya untuk mengevaluasi kinerja secara menyeluruh. Maka dari itu dibutuhkan peranan teknologi untuk membantu proses usaha, oleh karena itu, penulis memerlukan bantuan dari rapid miner (Gomez-Uribe & Hunt, 2015).

Dari uraian tersebut, maka bagaimana melakukan *classification process* pada data Netflix yang tersedia untuk mendapatkan hasil yang ingin penulis capai adalah memperkirakan genre film yang paling banyak ditonton/ ditayangkan oleh pengguna Netflix melalui aplikasi rapid miner. Dalam melakukan analisis penjualan dapat dilakukan dalam beberapa sudut pandang, salah satunya yaitu melakukan analisis terhadap film yang sering muncul dan ditonton oleh pelanggan (Prayitno & Rasim, 2018).

Dari semua ini penulis juga memerlukan data mining dari Netflix itu sendiri, data mining (Agarwal, 2014) adalah bidang persimpangan ilmu komputer dan

statistik yang digunakan untuk menemukan pola di bank informasi. Tujuan utama dari proses penambangan data adalah untuk mengekstraksi informasi yang relevant (Chakir dkk, 2020) serta berguna dari berkas data dan membentuknya menjadi struktur yang dapat dimengerti untuk penggunaan di masa depan. Ada berbagai proses dan teknik yang digunakan untuk melakukan penambangan data dengan sukses.

Setelah proses data mining sudah selesai, maka penulis dapat menentukan data data dari Netflix, seperti contohnya, data dan model dibalik rekomendasi Netflix (Amatriain, 2013) Sejak Netflix \$ 1 juta Hadiah, diumumkan pada tahun 2006, perusahaan kami telah dikenal memiliki personalisasi sebagai inti dari produk kami. Bahkan pada titik waktu itu, dataset yang kami rilis dianggap "besar", dan kami menggerakkan inovasi di bidang penelitian Penambangan Data.

TINJAUAN PUSTAKA

A. Data Mining

Data mining adalah suatu proses penambangan informasi penting dari suatu data. Informasi penting ini didapat dari suatu proses yang amat rumit seperti menggunakan artificial intelligence, teknik statistik, ilmu matematika, machine learning, dan lain sebagainya. Teknik-teknik rumit tersebut nantinya akan mengidentifikasi dan mengekstraksi informasi yang bermanfaat dari suatu database besar. Data mining telah menjadi disiplin ilmu yang dibangun dalam domain kecerdasan buatan (AI), dan rekayasa pengetahuan (KE). Data mining berakar pada machine learning dan statistika, tetapi merambah bidang lain dalam ilmu komputer dan ilmu lainnya seperti biologi, lingkungan, finansial, jaringan dan sebagainya.

Data mining telah mendapatkan begitu besar perhatian pada dekade terakhir sehubungan dengan perkembangan hardware yang menyediakan kemampuan komputasi luar biasa yang memungkinkan pengolahan

data besar. Tidak seperti kajian lain dalam AI dan KE, data mining dapat diperdebatkan sebagai sebuah aplikasi dibandingkan dengan sebuah teknologi, dengan demikian diharapkan akan menjadi topik yang hangat dibahas di masa mendatang, mengingat pertumbuhan data yang bersifat eksponensial. Paper ini memberikan kilas balik perjalanan sejarah data mining, keadaan saat ini dan beberapa pandangan dan perkembangan ke depan. Dan Asal usul data mining dapat dilihat kembali ke akhir tahun 1980-an pada saat istilah tersebut mulai digunakan, paling tidak dalam kalangan komunitas riset. Pada awalnya terdapat sedikit perdebatan tentang makna dan cakupan dari istilah tersebut dan sampai saat ini pertentangan tersebut masih terjadi. Dalam arti luas data mining dapat didefinisikan sebagai sekumpulan mekanisme dan teknik yang direalisasikan dalam perangkat lunak untuk mengekstrak informasi tersembunyi dari kumpulan data. Hal ini meningkatkan popularitas data mining yang secara bersamaan didukung oleh kemajuan teknologi, kemampuan CPU dan media yang menyimpan data dalam jumlah besar dan mengolahnya dalam waktu yang lebih cepat.

Data Mining diartikan sebagai menambang data atau upaya untuk menggali informasi yang berharga dan berguna pada database yang sangat besar (Agrawal et al., 1996). Hal terpenting dalam teknik data mining adalah aturan untuk menemukan pola frekuensi tinggi antar himpunan itemset yang disebut fungsi Association Rules (Aturan Asosiasi). Belakangan ini data mining telah diimplementasikan keberbagai bidang, diantaranya dalam bidang bisnis atau perdagangan, bidang pendidikan, dan telekomunikasi. Dibidang bisnis misalnya hasil implementasi data mining algoritma Apriori dapat membantu para pebisnis dalam pengambilan keputusan terhadap apa yang berhubungan dengan persediaan barang

B. Tahap – Tahap Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan

perantaraan knowledge base. Tahapan-tahapan tersebut, diantaranya :

1. Pembersihan data

Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. *Garbage in garbage out* (hanya sampah yang akan dihasilkan bila yang dimasukkan juga sampah) merupakan istilah yang sering dipakai untuk menggambarkan tahap ini. Pembersihan data juga akan mempengaruhi performansi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data

Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dsb. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. Dalam integrasi data ini juga perlu dilakukan transformasi dan pembersihan data karena seringkali data dari dua database berbeda tidak sama cara penulisannya atau bahkan data yang ada di satu database ternyata tidak ada di database lainnya.

3. Transformasi data

Beberapa teknik data mining membutuhkan format data yang khusus

sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan klastering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut binning. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik data mining yang dipakai. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik data mining tertentu yang tergantung pada tahapan ini.

4. Aplikasi data Teknik Mining

Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Ada beberapa teknik data mining yang sudah umum dipakai. Akan dibahas lebih jauh mengenai teknik-teknik yang ada di seksi berikutnya. Perlu diperhatikan bahwa ada kalanya teknik-teknik data mining umum yang tersedia di pasar tidak mencukupi untuk melaksanakan data mining di bidang tertentu atau untuk data tertentu. Sebagai contoh akhir-akhir ini dikembangkan berbagai teknik data mining baru untuk penerapan di bidang bioinformatika seperti analisa hasil microarray untuk mengidentifikasi DNA dan fungsinya.

5. Evaluasi pola yang di temukan

Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti : menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba teknik data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

6. Presentasi pola yang ditemukan untuk menghasilkan aksi

Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

C. Clustering

Clustering adalah salah satu tugas umum dari data mining, Pada tahap ini dilakukan pengelompokan data menggunakan kombinasi dua algoritma *clustering*, yaitu *hierarchical clustering* dan metode *K-means*. Dari algoritma *hierarchical clustering* ini digunakan untuk menentukan pusat *cluster*. Selanjutnya, pusat *cluster* yang diperoleh *hierarchical clustering* tersebut digunakan untuk proses pengelompokan data dengan menggunakan metode *K-means*

D. Rapidminer

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Teknologi Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (*Graphical User Interface*) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat *open source* dan dibuat dengan menggunakan program Java di bawah lisensi *GNU Public Licence* dan *Rapid Miner* dapat dijalankan di sistem operasi manapun. Dengan menggunakan Rapid Miner, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. Rapid Miner dikhususkan untuk penggunaan data mining. Model yang disediakan juga cukup banyak dan lengkap, seperti Model Bayesian, Modelling, Tree Induction, Neural Network dan lain-lain.

Banyak metode yang disediakan oleh Rapid Miner mulai dari klasifikasi, klustering, asosiasi dan lain-lain. Jika tidak ada model atau model algoritma yang tidak ada dalam Weka, pengguna boleh menambahkan modul lain, karena weka bersifat *open source*, jadi siapapun dapat ikut mengembangkan perangkat lunak ini.

METODE PENELITIAN

Adapun kerangka kerja dari penelitian ini yaitu sebagai berikut :

1. Mengkaji Bahan Literatur

Ini bertujuan untuk mendapatkan pemahaman yang dalam terhadap masalah yang akan diteliti dan metode yang digunakan sebagai solusi. Bahan literatur yang sangat penting publikasi karya ilmiah baik jurnal maupun *proceesing*, buku- buku terkait penggunaan software Rapidminer dan referensi lainnya.

2. Pengumpulan Data

Pengumpulan Data Pengumpulan data dilakukan di lokasi penelitian. Data yang dikumpul data yang berasal dari laporan hasil keuntungan dan melakukan proses wawancara sesuai dengan kebutuhan dari penerapan (Nofitri & Irawati, 2019).

3. Objek Penelitian

Data diambil pada kaggle dengan temanya adalah NetFlix Inc. Merupakan perusahaan multinational yang berasal dari Amerika dan bergerak di bidang entertainment. Netflix didirikan pada 29 Agustus 1998 di California oleh Reed Hastings dan Marc Randolph. Spesialisasi Netflix adalah penyedia layanan streaming media dan “video on demand online” dan DVD lewat surat. Pada tahun 2013, Netflix menambahkan produksi film maupun televisi, serta juga online distribution. Mirip langganan televisi berbayar (*cable tv*), Netflix bersih dari iklan, penonton tak perlu menunggu jadwal penayangan serial televisi, dan bisa menentukan sendiri konten yang ingin dinikmati (Hasan, 2017).

4. Ekstensi Data

Untuk melakukan proses *discoveri* pada RapidMiner perlu dilakukan ekstensi data. Ekstensi data yang dilakukan ialah melakukan ekstensi data dari format .csv ke format .xes. Format inilah yang nantinya digunakan untuk dilakukan pemodelan proses *discoveri* menggunakan aplikasi proses mining yaitu RapidMiner. Adapun dilakukan dengan menggunakan aplikasi ProM Import atau Disco. Aplikasi tersebut yang akan mengubah dari format .csv menjadi format .xes. (Rizqifaluthi & Yaqin, 2019)

5. Melakukan pengujian dengan Tools

RapidminerStudio 7.3. Pada tahap ini, hasil dari analisis dites kembali atau diuji lagi menggunakan sistem Data mining yang sudah ada. Sistem yang dipakai adalah Tools RapidminerStudio 7.3 yang mampu membaca data dari Microsoft Excel. (Fp-growth, 2017)

HASIL DAN PEMBAHASAN

Pada tahap awal, kami melakukan pengumpulan data netflix yang kami ambil dari situs kaggle. Data tersebut memiliki konten, yaitu :

1. Type

Type adalah merupakan data yang mengenai jenis tipe film dengan jenis TV Show atau Movie yang berada di platform streaming Netflix.

2. Title

Title adalah merupakan data mengenai judul film yang berada di konten dari platform streaming Netflix.

3. Director

Director adalah merupakan data mengenai yang mendirect sebuah film yang telah di sajikan.

4. Cast

Cast adalah merupakan data yang berisikan nama orang - orang yang mengisi peran dalam film tersebut.

5. Country

Country adalah merupakan data yang berisikan negara yang membuat film tersebut berasal.

6. *Date Added*

Date Added adalah merupakan data yang berisikan tentang tanggal film tersebut telah dirilis.

7. *Release Year*

Release Year adalah merupakan data yang berisikan tentang tahun film tersebut dirilis.

8. *Rating*

Rating adalah merupakan data yang berisikan tentang sistem klasifikasi usia untuk menonton film tersebut dengan jenis rating :TV-PG, TV-MA, TV-Y7-FV, TV-Y7, R, PG-13, NR , TV-14.

9. *Duration*

Duration adalah merupakan data yang berisikan durasi film tersebut yang dimana jenis nya : 90 min, 94 min, 1 Season, 110 min, 60 min.

10. *Listed in*

Listed in adalah merupakan data yang berisikan tentang daftar film yang di rekomendasikan untuk menonton.

11. *Description*

Description adalah merupakan data yang berisikan tentang deskripsi dari film tersebut.

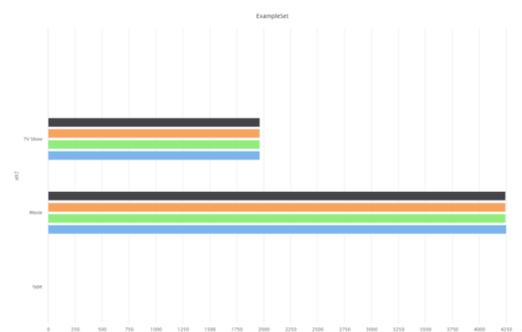
12. *Show Id*

Show Id adalah merupakan data yang berisikan tentang id yang terdapat di suatu film yang dapat mempermudah ketika mencari film tersebut di data yang ada.

A. Data selection

Data Selection merupakan proses meminimalkan jumlah data yang digunakan untuk proses mining dengan tetap merepresentasikan data aslinya.

Pada pembahasan ini, kami akan menjelaskan apa hasil dari penelitian yang sudah kami lakukan yaitu data Netflix dari segi banyaknya yang menonton jenis film tertentu terlebih dahulu. Di data kaggle kami, perusahaan Netflix yang merupakan sebuah platform yang menyediakan streaming video dan untuk menguji model prediksi sederhana , Netflix memiliki jumlah data 6,217 baris dan 12 kolom. Yang bisa di lihat pada Gambar 1.



Gambar 1. Tipe film yang sering dilihat

Data yang dianalisa dari perusahaan Netflix adalah mulai dari data jenis tipe film yang sering di lihat,tahun rilis,durasi dan rating film dari yang tertinggi hingga terendah dalam Netflix

Preprocessing Data

Preprocessing merupakan tahapan awal dalam mengolah data input sebelum memasuki proses tahapan utama, *Preprocessing* adalah salah satu tahapan penting untuk data mining. Data yang digunakan dalam proses penambangan tidak selalu dalam kondisi ideal untuk diproses.

Terkadang dalam data ada berbagai masalah yang dapat mengganggu hasil dari proses penambangan itu sendiri seperti nilai yang hilang, data yang berlebihan, outlier, atau format data yang tidak kompatibel dengan sistem. Oleh karena itu, untuk mengatasi masalah ini, diperlukan tahap *Preprocessing*.

Pertumbuhan data langganan Platform netflix yang telah tersedia kemudian dilakukan pengkodean menjadi kode "1" jika pertumbuhan suatu Tipe film yang sering di lihat lebih tinggi dibanding tahun sebelumnya, dan kode "0" jika lebih rendah dibanding tahun sebelumnya. Kode ini ditujukan untuk mengasumsikan data seperti penglihatan di suatu platform dimana data yang memiliki kode yang sama di tahun sama dianggap sebagai satu peningkatan data yang sama. Pada tahap ini juga dilakukan pengecekan untuk memastikan tidak adanya missing data dan duplikat data sehingga diperoleh data yang bersih. bisa dilihat pada Gambar 2

ID	Parent	Label	Leaf	Value
ast2	0	type (1)	Movie (4251)	Movie (4251), TV Show (965), [1 more]
ast3	0	ipad#4 J/C (1)	Turner (9)	Turner (9), Lindeus (2), [147 more]
ast4	0	Acad IggaMtr (1)	RA4 Ca er (988)	RA4 Campus, Jan Sahr (988), Marcus Raboy (14), [206 more]
ast5	0	AJRWaRkA Cen (1)	Dani Altonborogh (59)	Dani Altonborogh (59), Samud Hild (10), [411 more]
ast6	0	county (1)	United States (485)	United States (485), India (77), [553 more]
ast7	0	date_added (1)	January 1, 2020 (143)	January 1, 2020 (143), November 1, 2019 (104), [186 more]
ast8	0	release_year (1)	2018 (165)	2018 (165), 2017 (68), [1 more]
ast9	0	rating (1)	TV-14 (203)	TV-14 (203), TV-14 (196), [1 more]
ast10	0	duration (1)	1 Season (125)	1 Season (125), 2 Seasons (30), [199 more]
ast11	0	word_m (1)	Documentaries (30)	Documentaries (30), Stand-Up Comedy (254), [49 more]
ast12	0	description (1)	A suly Re (9)	A suly her Re (9), A rulle mndm (2), [321 more]

Gambar 2. Pengecekan Atribut

C. Data mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu database dengan melakukan penggalian pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basisdata. Tahap selanjutnya adalah menerapkan algoritma yang akan diujicobakan pada data Netflix.

Algoritma yang digunakan adalah algoritma Apriori. Algoritma akan diterapkan untuk melakukan identifikasi pola asosiasi pada data. Pengolahan data menggunakan tools RapidMiner. Running data dilakukan pada data masing-masing wilayah. Dan hasil data yang telah di running akan menjadi perwakilan data - data yang lain nya.

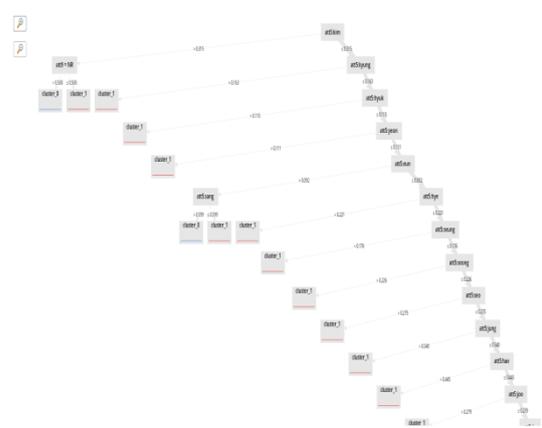
D. Decision Tree

Decision tree adalah sebuah diagram alir yang berbentuk seperti struktur pohon yang mana setiap internal node menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan output dari pegujian tersebut dan leaf node menyatakan kelas-kelas atau distribusi kelas.

Node yang paling atas disebut sebagai root node atau node akar. Sebuah root node akan memiliki beberapa edge keluar tetapi tidak memiliki edge masuk, internal node akan memiliki satu edge masuk dan beberapa edge keluar, sedangkan leaf node hanya akan

memiliki satu edge masuk tanpa memiliki edge keluar.

Decision tree digunakan untuk mengklasifikasikan suatu sampel data yang belum diketahui kelasnya ke dalam kelas-kelas yang sudah ada. Jalur pengujian data adalah pertama melalui root node dan terakhir adalah melalui leaf node yang akan menyimpulkan prediksi kelas bagi data tersebut. Atribut data harus berupa data kategorik, bila kontinu maka atribut harus didiskretisasi terlebih dahulu.



Gambar 3. Hasil dari Decision tree data Netflix

Berdasarkan pohon keputusan pada gambar 3, dapat ditarik kesimpulan dari klasifikasi bahwa terdapat 3 atribut yang berfungsi dalam *Decision tree* yaitu Type, Rating, Duration, dan release year. Maka proses penentuan prediksi hanya akan melihat 4 atribut itu berperan penting dalam analisa data netflix. Hasil klasifikasi pada data *training* adalah atribut charges sebagai *root* pada *decision tree* dan atribut lainnya sebagai *child* node.

E. Operator Apply Model

Operator Apply Model adalah Operator *Apply Model* digunakan untuk menerapkan model yang telah dilatih sebelumnya menggunakan data *training* pada *unlabeled data* (data *testing*). Tujuannya adalah untuk mendapatkan prediksi pada *unlabeled data* (data *testing*) yang belum memiliki label. Yang perlu diperhatikan adalah data *testing* harus memiliki urutan, jenis, maupun peran atribut yang sama dengan

DAFTAR PUSTAKA

- Agarwal, S. (2014). Data mining: Data mining concepts and techniques. *Proceedings International Conference on Machine Intelligence Research and Advancement, ICMIRA*.
<https://doi.org/10.1109/ICMIRA>.
- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, a I. (1996). Fast discovery of association rules. In *Advances in knowledge discovery and data mining*,5(2), 213–223.
- Amatriain, X. (2013). Big & personal: Data and models behind Netflix recommendations. *Proc. of 2nd Int. Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, BigMine 2013 - Held in Conj. with SIGKDD 2013 Conf*.
<https://doi.org/10.1145/2501221.2501222>
- Chakir, A., Chergui, M., dan Andry, A Smart Updater IT Governance Platform Based on Artificial Intelligence, *Advances in Science, Technology and Engineering Systems Journal* ,Vol. 5, No. 5, 47-53, 2020.
- Fp-growth, M. A. (2017). *Informatika* . 9(1), 15–25.
- Gomez-Uribe, C. A., & Hunt, N. (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*, 6(4).
<https://doi.org/10.1145/2843948>
- Hasan, V. A. (2017). Analisis Faktor-Faktor Yang Mempengaruhi Willingness To Subscribe: Telaah Pada Layanan Video On Demand Netflix. *Jurnal Manajemen*, 9(1),22–38.
<https://doi.org/10.31937/manajemen.v9i1.595>
- Kompas (2021), link:
<https://tekno.kompas.com/read/2016/01/07/13085347/Akhirnya.Masuk.Indonesia.Neflix.Itu.Apa>. Di akses pada tgl 22 Maret 2021.
- Nofitri, R., dan Irawati, N. (2019). *Intergrasi Metode Naive Bayes dan software Rapidminer dalam analisis hasil usaha perusahaan dagang perusahaan*.
Teknologi, J., Informasi, S., & No, V. I. 6(1), 35–42.
- Prayitno, M. H., & Rasim, R. (2018). Analisa Penjualan Produk Retail Dengan Metode Data Mining Asosiasi. *Jurnal Kajian Ilmiah*, 18(3), 231.
<https://doi.org/10.31599/jki.v18i3.273>
- Prediksi Rating Film Menggunakan Metode Naive Bayes. (2016). *Jurnal Teknik Elektro*.
- Rizqifaluthi, H., & Yaqin, M. A. (2019). Process Mining Akademik Sekolah menggunakan RapidMiner. *Matics*, 10(2),47-49.
<https://doi.org/10.18860/mat.v10i2.5158>
- Şeker dan Şadi, E. (2013). Rapid Miner. *Journal of Chemical Information and Modeling*.
<https://doi.org/10.1017/CBO9781107415324.004>