

**PENERAPAN METODE DECISION TREE
DENGAN ALGORITMA C4.5 UNTUK KLASIFIKASI
VIDEO GAMES RATING BY ESRB**

PROPOSAL TUGAS AKHIR



Diajukan Oleh :

Muhammad Bilal Alfayyadh

8020190044

Untuk Persyaratan Penelitian Dan Penulisan Tugas Akhir
Sebagai Akhir Proses Studi Strata 1

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DINAMIKA BANGSA
2022**

IDENTITAS PROPOSAL PENELITIAN

Judul Proposal : Penerapan Metode Decision Tree Dengan Algoritma C4.5
Untuk Klasifikasi Video Games Rating By ESRB

Program Studi : Teknik Informatika

Jenjang Pendidikan : Strata 1 (S1)

Peneliti :

- a. Nama Lengkap : Muhammad Bilal Alfayyadh
- b. NIM : 8020190044
- c. Jenis Kelamin : Laki-Laki
- d. Tempat/Tgl Lahir : Jambi/02 Juli 2001
- e. Alamat : Jl. Abdul Muis, RT.38, No.18,
Kec. Palmerah, Jerambah Bolong
- f. No. Telepon : 085357755969
- g. Email : bilalalfayyadh@gmail.com

ABSTRAK

Game merupakan jenis hiburan yang disukai oleh semua orang dari usia anak-anak, dewasa maupun tua. Selain digunakan untuk menghilangkan kepenatan dalam beraktivitas, sebuah *game* juga dapat berfungsi untuk melatih pola pikir seseorang. Tentunya *game* memiliki berbagai macam jenis, maka dari itu ada sebuah organisasi yang bernama ESRB yang melakukan penilaian terhadap *game* yang akan dimainkan oleh semua orang. Penilaian *game* tersebut memiliki dataset dalam format *xlsx* yang akan diuji oleh penulis dalam penelitian ini. Penulis akan melakukan pengujian dataset menggunakan tools berupa aplikasi WEKA dan RapidMiner yang nantinya akan membantu proses penelitian agar bisa mendapatkan hasil yang tepat dan sesuai.

Kata Kunci : Klasifikasi, *Decision Tree*, *Algoritma C4.5*, *Games Rating*

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan proposal skripsi yang berjudul “Penerapan Metode Decision Tree Dengan Algoritma C4.5 Untuk Klasifikasi Video Games Rating By ESRB” dengan sebaik-baiknya.

Penulis menyadari bahwa dalam penulisan proposal skripsi ini banyak mengalami kendala dan kekurangan, namun berkat bantuan, bimbingan, dan petunjuk dari banyak pihak sehingga kendala-kendala yang dihadapi tersebut dapat diatasi dan penulis dapat menyelesaikan proposal skripsi ini dengan baik.

Penulis berharap agar proposal skripsi ini dapat bermanfaat bagi kita semua. Akhir kata penulis mengucapkan terima kasih.

Jambi, September 2022

Penulis

DAFTAR ISI

COVER	i
IDENTITAS PROPOSAL PENELITIAN	ii
ABSTRAK	iii
KATA PENGANTAR.....	iv
DAFTAR ISI.....	v
BAB I PENDAHULUAN.....	1
1.1. LATAR BELAKANG	1
1.2. RUMUSAN MASALAH	2
1.3. BATASAN MASALAH	2
1.4. TUJUAN DAN MANFAAT PENELITIAN	2
1.4.1. Tujuan Penelitian	2
1.4.2. Manfaat Penelitian	2
BAB II LANDASAN TEORI	3
2.1. VIDEO GAMES RATING	3
2.2. DATA MINING.....	3
2.2.1. Tahapan Proses dalam Data Mining	4
2.2.2. Pengelompokkan Data Mining.....	7
2.3. KLASIFIKASI	9
2.4. DECISION TREE & ALGORITMA C4.5	9
2.5. PENELITIAN SEJENIS	11
BAB III METODOLOGI PENELITIAN	16
3.1. KERANGKA KERJA PENELITIAN	16
3.2. ANALISA TEKNIK ALGORITMA C4.5.....	18
3.3. METODE PENELITIAN.....	18
3.4. ALAT PENELITIAN.....	18
3.5. JADWAL PENELITIAN	19
DAFTAR PUSTAKA	
LAMPIRAN DATASET	

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG

Game merupakan jenis hiburan yang disukai oleh semua orang dari usia anak-anak, dewasa maupun tua. Selain digunakan untuk menghilangkan kepenatan dalam beraktivitas, sebuah *game* juga dapat berfungsi untuk melatih pola pikir seseorang untuk mencari solusi memecahkan suatu permasalahan yang ada di sebuah *game* [1]. Dahulu *game* dimainkan secara tradisional seperti permainan kartu, catur, ular tangga, petak umpet, dan lainnya. Seiring dengan berkembangnya teknologi, permainan tersebut dikembangkan ke dalam teknologi yang lebih modern. Sekarang banyak *game* baru yang memanfaatkan teknologi modern dalam pembuatannya dan penggunaannya.

Entertainment Software Rating Board (ESRB) adalah sebuah organisasi regulator mandiri yang menilai permainan video atau *video games*, panduan periklanan, prinsip privasi daring permainan video dan perangkat lunak hiburan lainnya di Kanada dan Amerika Serikat.

ESRB dibentuk pada tahun 1994 oleh Entertainment Software Association (sebelumnya Interactive Digital Software Association). Sejak tahun 2003, ESRB telah menilai 8.000 permainan yang diproduksi oleh 350 penerbit permainan video.

Salah satu alasan dibentuknya ESRB adalah karena kekerasan di dalam permainan seperti *Mortal Kombat* dan *Doom*, dan juga permainan video kontroversial lainnya yang mengandung hal-hal seksual.

Sehubungan dengan latar belakang diatas, penulis tertarik untuk melakukan penelitian yang berjudul ***“Penerapan Metode Decision Tree Dengan Algoritma C4.5 Untuk Klasifikasi Video Games Rating By ESRB”***.

1.2. RUMUSAN MASALAH

Berdasarkan dari latar belakang, maka didapat rumusan masalah sebagai berikut :

1. Bagaimana penerapan *Algoritma C4.5* dalam mengklasifikasi Rating Games By ESRB?
2. Seberapa besar tingkat akurasi *Algoritma C4.5* apabila digunakan untuk mengklasifikasi Rating Games By ESRB?

1.3. BATASAN MASALAH

Berdasarkan dari latar belakang, maka adapun batasan masalah dalam penelitian ini, yaitu:

1. Data yang diperlukan adalah dataset Rating Games By ESRB.
2. Variabel yang digunakan dalam penelitian ini adalah *Content Description* yang memberikan informasi mengenai konten didalam *game*.

1.4. TUJUAN DAN MANFAAT PENELITIAN

1.4.1. Tujuan Penelitian

Adapun tujuan dari penelitian yang akan dilakukan oleh penulis, yaitu :

1. Mengklasifikasi data Rating Games By ESRB
2. Mendapatkan akurasi yang tepat untuk melakukan klasifikasi Rating Games dengan menggunakan *Algoritma C4.5*.
3. Mengetahui berapa besar tingkat akurasi *Algoritma C4.5* dalam klasifikasi Rating Games By ESRB.

1.4.2. Manfaat Penelitian

Adapun beberapa manfaat yang akan didapat dalam melakukan penelitian ini, yaitu :

1. Memberikan kontribusi keilmuan pada penelitian bidang klasifikasi *data mining* khususnya untuk klasifikasi sebuah *rating* atau penilaian.
2. Menambah wawasan bagi penulis dalam pengklasifikasian dataset rating *video games*.

BAB II

LANDASAN TEORI

2.1. VIDEO GAMES RATING

Games Rating adalah sebuah sistem untuk mengelompokkan berbagai jenis *game* ke usia yang sesuai. Biasanya, rating akan menilai konten yang ada didalam *game* itu sendiri. Misalnya, jika sebuah *game* berisi tindakan kekerasan, maka game tersebut akan dimasukkan ke kategori *game* dewasa.

2.2. DATA MINING

Data mining sering juga disebut *Knowledge Discovery in Database*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan [2].

Menurut Turban et al. [3], “*data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam *database* besar”.

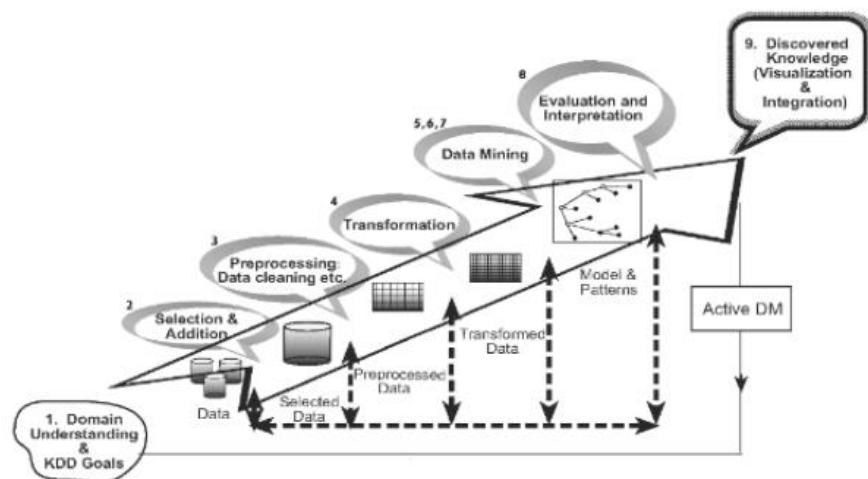
Adapun definisi lainnya, seperti Pramudiono [4] mendefinisikan “*data mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya”.

Berdasarkan definisi – definisi yang telah disampaikan, hal penting yang terkait dengan *data mining* adalah :

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada dalam sebuah data yang bervolume besar.
2. Data yang akan di proses berupa data yang sangat besar. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

2.2.1. Tahapan Proses dalam *Data Mining*

Ada beberapa tahapan dalam proses *data mining*. Diagram dibawah ini menggambarkan beberapa tahap/proses yang berlangsung dalam *data mining* [5].



Gambar 2.1. Fase – Fase dalam *Data Mining*

1. **Pembentukan pemahaman domain aplikasi.** Pada tahap ini menentukan tujuan dari end-user dan bagian terkait dimana KDD dilakukan. Mengembangkan pemahaman tentang domain aplikasi ini adalah awal langkah persiapan. Ini mempersiapkan adegan untuk memahami apa yang harus dilakukan dengan banyak keputusan (tentang transformasi, algoritma, representasi, dll). Orang-orang yang bertanggung jawab atas proyek KDD perlu memahami dan menentukan tujuan dari pengguna akhir dan lingkungan di mana pengetahuan penemuan proses akan berlangsung

(termasuk pengetahuan awal yang relevan). Sebagai KDD hasil proses, mungkin ada revisi dan perbaikan dari langkah ini. Memiliki memahami tujuan KDD, preprocessing data dimulai, sebagaimana didefinisikan dalam berikutnya tiga langkah (perhatikan bahwa beberapa metode di sini mirip dengan algoritma Data Mining, tetapi digunakan dalam konteks preprocessing).

2. **Memilih dan menciptakan** satu data set untuk mendukung proses penemuan knowledge akan dilakukan. Penentuan data yang akan digunakan untuk proses KDD dilakukan pada tahap ini. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data untuk KDD ke dalam sebuah data set, termasuk atribut yang diperlukan dalam proses KDD. Terdapat interaktif dan iteratif dari KDD tersebut. Dimulai dengan data yang tersedia baik mengatur dan kemudian mengembang dan mengamati efeknya dalam KDD.
3. **Preprocessing dan cleansing.** Dalam tahap ini kehandalan data ditingkatkan. Termasuk data clearing, seperti menangani data yang tidak lengkap, menghilangkan gangguan atau outlier. Termasuk menggunakan metode statistik yang kompleks, atau melakukan penambangan spesifik data dengan algoritma dalam KDD.
4. **Transformasi data.** Pada tahap ini, generasi data yang lebih baik untuk data mining dipersiapkan dan dikembangkan, membuat data menjadi lebih baik menggunakan metode reduksi dimensi dan transformasi atribut. Sebagai contoh, dalam pemeriksaan medis, hasil bagi atribut mungkin sering menjadi faktor yang paling penting, dan tidak satu persatu. Di pemasaran, kita mungkin perlu mempertimbangkan efek di luar kendali kita serta upaya dan temporal isu (seperti mempelajari pengaruh akumulasi iklan). Namun, bahkan jika kita tidak menggunakan transformasi yang tepat di awal, kita dapat memperoleh efek mengejutkan bahwa petunjuk kepada kita tentang transformasi diperlukan (Pada iterasi berikutnya). Dengan demikian proses KDD mencerminkan kepada dirinya sendiri dan menyebabkan pemahaman tentang transformasi yang dibutuhkan (seperti pengetahuan ringkas dari sebuah ahli dalam bidang tertentu mengenai indikator terkemuka kunci).

5. **Memilih tugas data mining** yang cocok. Pada tahap ini ditentukan tipe data mining yang akan digunakan, apakah klasifikasi, regresi, atau clustering, tergantung pada tujuan KDD dan tahap sebelumnya.
6. **Memilih algoritma data mining.** Pemilihan algoritma yang paling tepat untuk menemukan pola dilakukan pada tahap ini. Ada dua tujuan utama dalam Data Mining: prediksi dan deskripsi. Prediksi sering disebut sebagai Supervised Data Mining, sementara deskriptif Data Mining meliputi aspek-aspek unsupervised dan visualisasi Data Mining. Sebagian besar data teknik pertambangan didasarkan pada pembelajaran induktif, di mana model yang dibangun secara eksplisit maupun implisit oleh generalisasi dari jumlah yang memadai pelatihan data training. Asumsi yang mendasari pendekatan induktif adalah bahwa model terlatih ini berlaku untuk kasus masa depan. Strategi ini juga memperhitungkan tingkat metalearning untuk set tertentu dari data yang tersedia.
7. **Penggunaan algoritma data mining.** Pada tahap ini dilakukan implementasi dari algoritma data mining yang telah ditentukan pada tahap sebelumnya. Misalnya dengan menseting parameter kontrol algoritma, seperti jumlah minimum kasus dalam daun tunggal dari pohon keputusan.
8. **Evaluasi,** pada tahap ini dilakukan evaluasi dan penerjemahan dari pola yang diperoleh, sehubungan dengan tujuan yang ditetapkan pada langkah pertama. Langkah ini berfokus pada komprehensibilitas dan kegunaan dari model induksi. Pada langkah ini pengetahuan ditemukan juga terdokumentasi untuk penggunaan lebih lanjut. Langkah terakhir adalah penggunaan dan umpan balik secara keseluruhan pada pola dan hasil penemuan diperoleh dengan Data Mining.
9. **Penggunaan pengetahuan yang ditemukan** yakni memasukkan pengetahuan ke dalam sistem lain untuk ditindaklanjuti. Pengetahuan menjadi aktif dalam arti bahwa kita dapat membuat perubahan ke sistem dan mengukur dampak. Sebenarnya keberhasilan langkah ini menentukan efektivitas proses KDD secara keseluruhan. Ada banyak tantangan dalam langkah ini, kehilangan "kondisi laboratorium". Misalnya, pengetahuan itu

ditemukan dari sebuah snapshot statis tertentu (biasanya sampel) dari data, tapi sekarang data menjadi dinamis.

2.2.2. Pengelompokan *Data Mining*

Pengelompokan *data mining* dapat diklasifikasikan berdasarkan fungsi yang dilakukan atau berdasarkan jenis aplikasi yang menggunakannya, yaitu [4] :

1. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

Contoh lain klasifikasi dalam bisnis dan penelitian adalah :

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
- b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosa penyakit seorang pasien untuk mendapatkan termasuk kategori apa.

2. Pengklusteran (*Clustering*)

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

Contoh pengklusteran dalam bisnis dan penelitian adalah :

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial dalam baik dan mencurigakan.

3. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Contoh asosiasi dalam bisnis dan penelitian adalah :

- a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran *upgrade* layanan yang diberikan.
- b. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli bersamaan.

4. Deskripsi

Terkadang analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

5. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level

sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

6. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang.

2.3. KLASIFIKASI

Klasifikasi adalah tugas pembelajaran sebuah fungsi target yang memetakan setiap himpunan atribut ke salah satu label *class* yang telah didefinisikan sebelumnya [6].

Klasifikasi dapat juga diartikan suatu proses untuk menemukan suatu model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan dapat menggunakan model untuk memprediksi kelas objek yang label *class*-nya tidak diketahui.

2.4. DECISION TREE & ALGORITMA C4.5

Pohon keputusan atau *decision tree* merupakan teknik untuk mengidentifikasi tindakan yang menghasilkan sebuah keputusan dan konsekuensi sebagai garis antara node (lingkaran dalam diagram).

Menurut Surjeet Kumar et al. [7], “*decision tree is a flow-chart-like tree structure, where each internal node is denoted by rectangles, and leaf nodes are denoted by ovals*”.

Menurut Michael J. A. et al. [8], “*decisison tree* merupakan struktur yang dapat digunakan untuk membagi fakta yang besar menjadi pohon keputusan yang mempresentasikan aturan, kemudian aturan tersebut dapat dengan mudah untuk diinterpretasi oleh manusia. Manfaat lain dari pohon keputusan adalah untuk mengeksplorasi data, mencari keterkaitan tersembunyi antara sejumlah variabel input dengan variabel target. Pohon keputusan biasanya menunjukkan data dalam bentuk tabel dengan atribut dan record. Atribut menunjukkan suatu parameter yang dibuat sebagai kriteria dalam pembentukan *tree*.”

Algoritma C4.5 digunakan untuk membentuk pohon keputusan (Decision Tree). Algoritma C4.5 merupakan pengembangan dari algoritma ID3, proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule. [6][9]

Algoritma C4.5 menggunakan konsep information gain atau entropy reduction untuk memilih pembagian yang optimal [9]. Tahapan dalam membuat pohon keputusan dengan algoritma C4.5, yaitu [10] :

1. Mempersiapkan data training, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai gain yang tertinggi dari masing-masing atribut atau berdasarkan nilai index entropy terendah. Sebelumnya dihitung terlebih dahulu nilai index entropy, dengan rumus :

$$Entropy(i) = - \sum_{j=1}^m f(i,j). \log_2 f[(i,j)]$$

3. Hitung nilai gain dengan rumus :

$$Entropy\ split = \sum_{i=1}^p \binom{n1}{n} \cdot IE(i)$$

4. Ulangi langkah ke-2 hingga semua record terpartisi. Proses partisi pohon keputusan akan berhenti disaat :
 - a. Semua tupel dalam record dalam simpul N mendapat kelas yang sama
 - b. Tidak ada atribut dalam record yang dipartisi lagi
 - c. Tidak ada record di dalam cabang yang kosong

2.5. PENELITIAN SEJENIS

Tabel 2.1. Penelitian Sejenis

No.	Judul, Penulis dan Tahun	Metode	Atribut	Akurasi
1	“Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Pengaruh Negatif Game Online Bagi Remaja Milenial”. Nungky Asmiati, Fatmawati. 2020 [11]	<i>Algoritma Naïve Bayes</i>	Nama Lengkap, Jenis Kelamin, Usia, Jumlah Hari, Durasi Waktu, Klasifikasi	Dari 100 data sample pengguna game online masih dikategorikan “normal” dengan akurasi keseluruhan 89.00%, recall kecanduan 77.27% dan recall normal 92.31%, precision kecanduan 73.91% dan precision normal 93.51%.
2	“Analisis Sentimen Untuk Mengetahui Kesan Player Game Mobile Legends Menggunakan Naïve Bayes Classifier”.	<i>Naïve Bayes Classifier</i>	Kata Baku, Bahasa Inggris, Slang	<i>Algoritma Naive Bayes Classifier</i> menghasilkan data paling banyak yaitu data komentar/opini negative dengan menggunakan

	Daurat Sinaga, Cahaya Jatmoko. 2020 [12]			Confusion Matrix mendapatkan akurasi sebesar 80%.
3	“Analisa Klasifikasi Genre Game Pc Terpopuler”. Muhammad Rivaldy Hisham, Jumiliono Pratama, Luky Andito, Andy Kho, Hendry Wijaya. 2022 [13]	<i>Random Forest Classifier, Gradient Boosting Decision</i>	Ranking, Nama, Platform, Tahun, Genre, Penerbit, Penjualan Amerika Utara, Penjualan Uni Eropa, Penjualan Jepang, Penjualan Yang Lain, Penjualan Internasional	Penggunaan algoritma <i>Random Forest</i> dalam pengklasifikasian game populer sudah tepat dengan tingkat akurasi 100%. Pada <i>classification report gradient boosting</i> tingkat keakurasian pada algoritma sebesar 94.9%.
4	“Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku”. Mawadatul Maulidah, Windu Gata, Rizki Aulianita, Cucu Ika	<i>Decision Tree, Algoritma C4.5</i>	bookID, title, author, average_rating, isbn, isbn13, language_code, num_pages, rating_count, text_review _count, publication_date, publisher	Hasil dari evaluasi dan validasi, diketahui bahwa Decision Tree memiliki akurasi yang paling tinggi sebesar 99,95%, precision 100%, recall 96%, f1- score 98%.

	Agustyaningrum. 2020 [14]			
5	“Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest”. Green Arther Sandag. 2020 [15]	<i>Algoritma Random Forest</i>	Id, Track_Name, Size_Bytes, Currency, Price, Rating_Count_Tot, Rating_Count_Ver, User_Rating, User_Rating_Ver, Ver, Cont_Rating, Prime_Genre, Sup_Devices.Num, Ipadsc_Urls.Num, Lang.Num, Vpp_Lic, App_Desc	Hasil dari 10 Fold Cross-Validation didapati Accuracy 86.08%, Recall 84.56%, Precision 84.38%, dan RMSE 0.316 dalam Memprediksi Rating pada Apple’s Appstore.
6	“Genre E-Sport Gaming Tournament Classification Using Machine Learning Technique Based On Decision Tree, Naïve Bayes, And Random Forest Algorithm”. Arif Rinaldi Dikananda, Irfan Ali, Fathurrohman, Ade Rizki Rinaldi, Iin. 2020 [16]	<i>Decision Tree, Naïve Bayes, Random Forest Algorithm</i>	Id, Gender, Total Income, Internet Benefit, General Pay	The result from this research is <i>Random Forest</i> Accuration 60%.

7	“Klasifikasi Ulasan Aplikasi Jenius pada Google Play Store Menggunakan Algoritma Naive Bayes”. Raihan Adyatma Subagja, Yuni Widiastiwi, Nurul Chamidah. 2021 [17]	<i>Algoritma Naïve Bayes</i>	Aplikasi, Bagus, Banyak, Lengkap, Login, Mudah, Fitur, Lama, Lambat, Sedikit, Susah, Transaksi	Pengujian <i>Naive Bayes Classifier</i> menghasilkan akurasi sebesar 57%, hasil precision dan recall sebesar 78% dan 87%.
8	“Implementasi Algoritma Naïve Bayes Dalam Penentuan Rating Buku”. Muthia Anggraini, Rizki Ayuning Tyas, Ismi Ana Sulasiyah, Qurrotul Aini. 2020 [18]	<i>Algoritma Naïve Bayes</i>	bookID, Title, Author, ISBN, isbn13, language_code, num_pages, rating_counts, text_review_count, publication_date, Publisher	Hasil penelitian menentukan rating buku menggunakan metode <i>Naïve Bayes</i> memiliki accuracy 66.98%, precision 74.47% dan recall 62.47%.
9	“Prediction Of The Digital Game Rating Systems Based On The ESRB”. Khaled Mohammad Alomari, Ahmad Qasim Alhamad, Hisham O. Mbaidin, Said Salloum. 2019 [19]	<i>Generalized Linear, Decision Tree, Deep Learning, Gradient Boosted, Random Forest, Naïve Bayes</i>	Game Title, Rate, 38 Content Descriptors	Generalized Linear Models (GLMs) achieved a best accuracy 90.27%, Decision Tree 89.78%, Deep Learning 89.78%, Gradient Boosted 89.05%, Random Forest 83.70%, Naïve Bayes 77.86%.

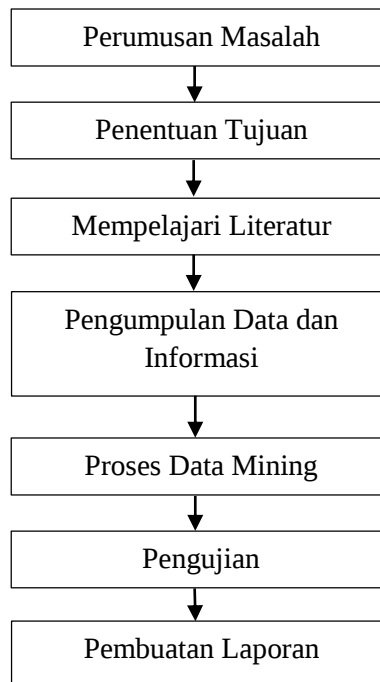
10	“Interpretable Models For The Potentially Harmful Content In Video Games Based On Game Rating Predictions”. Feng Zhipeng, Hamdan Gani. 2021 [20]	<i>Random Forest</i>	Game Title, Rate, 38 Content Descriptors	The interpretability analysis, resulting in an accuracy of 84.20%, a precision of 85.00%, a recall of 84.00%, and F1-score of 84.00%.
----	--	-----------------------------	---	--

BAB III

METODOLOGI PENELITIAN

3.1. KERANGKA KERJA PENELITIAN

Untuk membantu dalam penyusunan penelitian ini, maka perlu adanya kerangka kerja (*framework*) yang jelas tahapan-tahapannya. Kerangka kerja ini merupakan langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Adapun kerangka kerja penelitian yang akan digunakan adalah sebagai berikut :



Gambar 3.1. Tahapan Penelitian

Berdasarkan kerangka kerja penelitian yang telah di gambar di atas, maka dapat di uraikan pembahasan masing masing tahapan dalam penelitian adalah sebagai berikut :

1. Perumusan Masalah

Masalah yang dirumuskan dalam penelitian ini adalah bagaimana penerapan metode *Decision Tree* dengan *Algoritma C4.5* dalam klasifikasi Video Games Rating By ESRB.

2. Penentuan Tujuan

Tujuan yang akan dicapai dalam penelitian ini adalah ingin mengetahui berapa besar tingkat akurasi *Algoritma C4.5* dalam klasifikasi Rating Games By ESRB.

3. Mempelajari Literatur

Mempelajari literatur-literatur yang dapat mencapai tujuan penelitian, literatur-literatur bersumber dari buku-buku perpustakaan dan jaringan internet. Literatur-literatur yang digunakan nanti dilampirkan dalam daftar pustaka.

4. Pengumpulan Data dan Informasi

Dalam pengumpulan data, penulis mendapatkan dataset online yang terdapat pada sebuah website di internet.

5. Proses Data Mining

Data Mining adalah proses pengekstrasian *knowledge* yang tersimpan dalam dataset bervolume besar. Untuk mendapatkan *knowledge* dalam dataset digunakanlah *Algoritma C4.5*.

6. Pengujian

Pada tahap ini dilakukan pengujian dari hasil yang didapat dari tahap sebelumnya sebagai pedoman untuk mendapatkan hasil klasifikasi Rating Games By ESRB.

7. Pembuatan Laporan

Pada tahap ini dilakukan pembuatan laporan, membuat hasil akhir dari suatu kegiatan penelitian berdasarkan data dan fakta yang telah diamati pada saat meneliti.

3.2. ANALISA TEKNIK ALGORITMA C4.5

Pada bagian ini data dan informasi yang diperoleh dan diproses menggunakan Metode *Decision Tree* dengan *Algoritma C4.5* untuk mendapatkan hasil yang sesuai.

3.3. METODE PENELITIAN

Metode pengumpulan data yang digunakan dalam penelitian adalah mencari dan mengambil *dataset* online dari sebuah website yang bernama “*www.kaggle.com*” untuk dijadikan pengujian dalam penelitian ini. Pengumpulan data dilakukan untuk memperoleh hasil dan informasi yang dibutuhkan dalam hal mencapai tujuan penelitian.

3.4. ALAT PENELITIAN

Untuk membantu proses penelitian ini, penulis menggunakan beberapa alat/piranti yang digunakan untuk melakukan pengolahan data/bahan penelitian, yaitu:

1. Hardware, dengan spesifikasi sebagai berikut:
 - a. Laptop, dengan processor Intel(R) Core-i5 (2.40 GHz)
 - b. RAM : 8.00 GB
 - c. SSHD : 1 TB
2. Software, dengan keterangan sebagai berikut:
 - a. OS Windows 11 (64 bit)
 - b. Microsoft Excel & Word 2016
 - c. WEKA
 - d. RapidMiner Studio Professional 7.1.001

3.5. JADWAL PENELITIAN

No	Jenis Kegiatan	Bulan																							
		Sept				Okt				Nov				Des				Jan				Feb			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Perumusan Masalah																								
2	Penentuan Tujuan																								
3	Mempelajari Literatur																								
4	Pengumpulan Data & Informasi																								
5	Proses Data Mining																								
6	Pengujian																								
7	Pembuatan Laporan																								

DAFTAR PUSTAKA

- [1] S. Yogesh, S. Ratna, and T. Anjana, "Immediate and long-term effects of meditation on acute stress reactivity, cognitive functions, and intelligence," *Altern. Ther. Health Med.*, vol. 18, no. 6, pp. 46–53, 2012.
- [2] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [3] E. Turban, J. E. Aronson, and T. P. Liang, "Decision Support Systems and Intelligent System,(Sistem Pendukung Keputusan dan Sistem Cerdas) Ed. 7. Jld. 2," 2005.
- [4] E. T. L. Kusri and E. Taufiq, "Algoritma data mining," *Yogyakarta Andi Offset*, 2009.
- [5] L. Rokach and O. Maimon, "Clustering methods," in *Data mining and knowledge discovery handbook*, Springer, 2005, pp. 321–352.
- [6] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," *Data Min. Concepts Tech.*, 2012, doi: 10.1016/C2009-0-61819-5.
- [7] S. K. Yadav and S. Pal, "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification," vol. 2, no. 2, pp. 51–56, 2012, [Online]. Available: <http://arxiv.org/abs/1203.3832>
- [8] M. J. A. Berry and G. S. Linoff, *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 2004.
- [9] D. T. Larose, D. Ph, C. D. Larose, and D. Ph, "Solutions to Chapter 1," *Algebr. Through Pract.*, pp. 29–50, 2012, doi: 10.1017/cbo9781139168274.007.
- [10] F. Gorunescu, *Data Mining: Concepts, models and techniques*, vol. 12. Springer Science & Business Media, 2011.
- [11] N. Asmiati, S. Informasi, S. Tinggi, M. Informatika, D. Komputer, and N. Mandiri, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi

Pengaruh Negatif Game Online Bagi Remaja Milenial (Application of the Naive Bayes Algorithm to Classify the Negative Effects of Online Games on Millennial Adolescents),” vol. 2, no. 3, pp. 141–149, 2020.

- [12] D. Sinaga and C. Jatmoko, “Analisis Sentimen Untuk Mengetahui Kesan Player Game Mobile Legends Menggunakan Naïve Bayes Classifier,” *Semnas LPPM UMP*, vol. V, pp. 540–547, 2020.
- [13] M. R. Hisham, J. Pratama, L. Andito, A. Kho, and H. Wijaya, “Analisa Klasifikasi Genre Game PC Terpopuler,” *J. Inf. Syst. Hosp. Technol.*, vol. 4, no. 01, pp. 27–31, 2022, doi: 10.37823/insight.v4i01.145.
- [14] M. Maulidah, Windu Gata, Rizki Aulianita, and Cucu Ika Agustyaningrum, “Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku,” *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 89–96, 2020, doi: 10.51903/e-bisnis.v13i2.251.
- [15] G. A. Sandag, “Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest,” *CogITO Smart J.*, vol. 6, no. 2, p. 167, 2020, doi: 10.31154/cogito.v6i2.270.167-178.
- [16] A. R. Dikananda, I. Ali, Fathurrohman, R. Ade Rinaldi, and Iin, “Genre e-sport gaming tournament classification using machine learning technique based on decision tree, Naïve Bayes, and random forest algorithm,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012037, 2021, doi: 10.1088/1757-899x/1088/1/012037.
- [17] R. Adyatma Subagja, Y. Widiastiwi, and N. Chamidah, “Klasifikasi Ulasan Aplikasi Jenius pada Google Play Store Menggunakan Algoritma Naive Bayes,” *Inform. J. Ilmu Komput.*, vol. 17, no. 3, p. 197, 2021, doi: 10.52958/iftk.v17i3.3652.
- [18] R. A. Tyas, M. Anggraini, I. A. Sulasiyah, and Q. Aini, “Implementasi Algoritma Naïve Bayes Dalam Penentuan Rating Buku,” *Sistemasi*, vol. 9, no. 3, p. 557, 2020, doi: 10.32520/stmsi.v9i3.915.
- [19] K. M. Alomari, A. Q. Alhamad, H. O. Mbaidin, and S. Salloum,

“Prediction of the digital game rating systems based on the ESRB
[Predicción de los sistemas de clasificación de juegos digitales basados en
el sistema de calificación ESRB],” *Opcion*, vol. 35, no. Special Issue 19,
pp. 1368–1393, 2019, [Online]. Available:
[https://www.scopus.com/inward/record.uri?eid=2-s2.0-
85068870131&partnerID=40&md5=646bdf7f8ad25726e514d0a48fb1818a](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85068870131&partnerID=40&md5=646bdf7f8ad25726e514d0a48fb1818a)

- [20] F. Zhipeng and H. Gani, “Interpretable Models for the Potentially Harmful
Content in Video Games Based on Game Rating Predictions,” *Appl. Artif.
Intell.*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2021.2008148.

LAMPIRAN DATASET

Title	Console	Alcohol Reference	Animated Blood	Blood			Use of Alcohol	Use of Drugs and Alcohol	ESRB Rating
Monster Jam Steel Titans 2	1	0	0	0			0	0	E
Subnautica: Below Zero	1	0	1	0			0	0	ET
NIER REPLICANT VER.1.22474487139&€!	1	0	0	1			0	0	M
Jamestown+	0	0	0	0			0	0	ET
Neptunia Virtual Stars	0	0	0	0			0	0	T
Monster Energy Supercross - The Official Vi	1	0	0	0			0	0	E
Monochrome Order	0	0	0	1			0	0	T
Blightbound	1	0	0	0			0	0	T
Maquette	0	1	0	0			0	0	ET
FATAL FURY&,c BATTLE ARCHIVES VOL. 2	0	0	0	0			0	0	T
NBA LIVE 20	1	0	0	0			0	0	E
Golden Force	1	0	0	1			0	0	M
Dungreed	0	1	0	0			0	0	ET
The End is Nigh	0	0	0	0			0	0	M
Fallout 76: Steel Dawn	1	0	0	0			0	0	M
Red Dead Online	1	0	0	0			0	0	M
Override 2: Super Mech League &€" Ultram	1	0	0	0			0	0	T
Turrican Flashback	0	0	0	0			0	0	ET
Sniper Ghost Warrior Contracts 2	1	0	0	0			0	0	M
Surviving the Aftermath	0	0	0	0			0	0	T
Asdivine Menace	0	0	0	0			0	0	T
West of Dead	0	0	0	0			0	0	T
FIVE NIGHTS AT FREDDY'S: HELP WANTED	0	0	0	0			0	0	T
Ruined King: A League of Legends Story	1	1	0	1			0	0	T
Tanuki Justice	0	0	0	0			0	0	ET
Bibi & Tina at the Horse Farm	0	0	0	0			0	0	E
Bugsntax	0	0	1	0			0	0	ET
Bus Simulator 21	1	0	0	0			0	0	E
Five Nights at Freddy's: Core Collection	1	0	0	0			0	0	T
Afterparty	0	0	0	1			0	0	M
EA SPORTS&,c FIFA 21 &€%DITION NXT LVL	1	0	0	0			0	0	E
EA SPORTS&,c FIFA 21 NXT LEVEL EDITION	1	0	0	0			0	0	E
EA SPORTS&,c MADDEN NFL 21 &€%DITION	1	0	0	0			0	0	E
EA SPORTS&,c MADDEN NFL 21 NXT LEVEL	1	0	0	0			0	0	E
My Universe - Pet Clinic Cats and Dogs	0	0	0	0			0	0	E
Empire of Sin	1	1	0	1			0	0	M
Worms Rumble	0	0	0	0			0	0	ET
EarthNight	0	0	0	0			0	0	ET
METAL SLUG ANTHOLOGY&,c	0	0	0	1			0	0	T
NHL 94 Rewind	1	0	0	0			0	0	E
Story of a Gladiator	0	0	0	0			0	0	M
The TakeOver	0	0	0	0			0	0	T
Space Invaders Forever	0	0	0	0			0	0	E
BALAN WONDERWORLD	1	0	0	0			0	0	ET
HITMAN 3	1	0	0	1			0	0	M
Bubble Bobble 4 Friends: The Baron is Back	0	0	0	0			0	0	E
Fortnite	1	0	0	0			0	0	T
Fortnite - The Last Laugh Bundle	1	0	0	0			0	0	T
Who Wants to Be a Millionaire?	1	1	0	0			0	0	E
Obduction	0	1	0	0			0	0	ET
.....									
.....									
SPARC	0	0	0	0			0	0	E
Still Time	0	0	0	1			0	0	T
Surf World Series	1	0	0	0			0	0	E

Keterangan :

- 1 Title (Judul Game)
- 1 ESRB Rating (Kategori Game)
- 32 Content Descriptors (Kolom Atribut)

Sumber Dataset :

<https://www.kaggle.com/datasets/imohtn/video-games-rating-by-esrb>