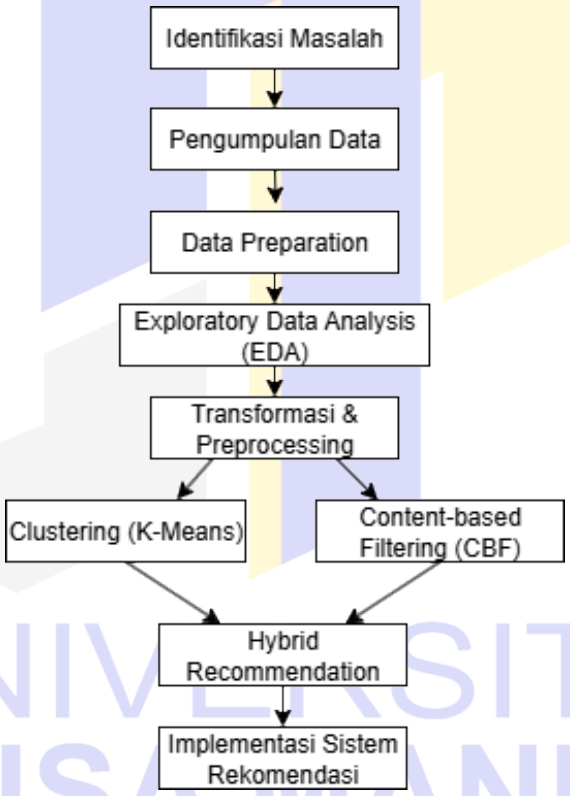
BAB III

METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian terapan (*applied research*) yang memanfaatkan metode *machine learning* untuk membangun sistem rekomendasi *game* berbasis web. Pendekatan yang digunakan adalah kuantitatif dengan metode eksperimen, karena melibatkan pengolahan data numerik dari API RAWG yang dianalisis menggunakan algoritma *machine learning*.

3.1 Konsep Penelitian



Sumber: Hasil penelitian

Gambar III. 1 Tahapan penelitian

Dalam kerangka berfikir yang disusun oleh penulis, diuraikan pada gambar III.1 sebagai berikut:

a. Identifikasi Masalah

Pada tahap ini peneliti mengidentifikasi masalah yang menjadi dasar perlunya sistem rekomendasi *game*. Seiring dengan pertumbuhan industri *game* yang semakin pesat, pengguna seringkali mengalami kesulitan dalam menemukan *game* yang sesuai dengan preferensi mereka. Banyaknya jumlah *game* yang tersedia di berbagai platform membuat proses pemilihan *game* menjadi kurang efisien dan memakan waktu. Selain itu, penelitian sebelumnya umumnya hanya memanfaatkan satu metode rekomendasi, seperti *collaborative filtering* atau *content-based filtering*, tanpa menggabungkan keduanya dengan teknik *clustering*. Oleh karena itu, penelitian ini memfokuskan diri pada pembangunan sistem rekomendasi berbasis *hybrid*, yaitu menggabungkan algoritma *K-Means* dan *content-based filtering* untuk meningkatkan kualitas rekomendasi yang dihasilkan.

b. Pengumpulan Data

Setelah permasalahan teridentifikasi, langkah berikutnya adalah pengumpulan data. Data yang digunakan dalam penelitian ini berasal dari RAWG *Video Games Database* API. Dari sumber ini, peneliti mengambil sebanyak 1.000 data *game* dengan atribut yang dianggap penting dalam menentukan relevansi rekomendasi, yaitu:

- *Nama game*
- *Rating*
- *Genre*
- *Platform*
- Kategori usia (ESRB)

Proses pengambilan data dilakukan dengan menggunakan bahasa

pemrograman Python di *Google Colaboratory*. Data hasil ekstraksi kemudian disimpan dalam format *Comma Separated Values* (CSV) agar mudah diolah pada tahap berikutnya.

c. Data *Preparation*

Tahap data *preparation* dilakukan untuk memastikan dataset dalam kondisi bersih dan siap digunakan dalam pemodelan. Proses ini melibatkan:

1) Pemeriksaan kelengkapan data, yaitu memastikan semua atribut yang diperlukan tersedia untuk setiap entri.
2) Identifikasi dan penghapusan data duplikat, agar tidak terjadi bias dalam analisis.
3) Seleksi atribut relevan, dengan hanya mempertahankan atribut yang benar-benar berpengaruh terhadap proses rekomendasi.

Dengan demikian, dataset yang digunakan pada penelitian ini benar-benar terfokus pada fitur penting yang mendukung proses rekomendasi.

d. *Exploratory Data Analysis* (EDA)

Tahap ini bertujuan memahami karakteristik dataset secara lebih mendalam melalui eksplorasi pola distribusi data. Analisis yang dilakukan meliputi:

1) Distribusi *rating game*, yang menunjukkan sebagian besar *game* memiliki rating pada rentang 3.0–4.5.
2) Distribusi kategori usia (ESRB), di mana mayoritas *game* termasuk ke dalam kategori *Mature*.
3) Distribusi genre *game*, yang memperlihatkan dominasi *genre Action, Adventure, dan Indie.*
4) Distribusi *platform game*, dengan PC sebagai *platform* dominan,

diikuti oleh *Android* dan *PlayStation* 4.

Hasil eksplorasi data divisualisasikan dalam bentuk histogram, diagram batang, dan grafik distribusi. EDA ini penting karena memberikan gambaran umum dataset serta mendukung pengambilan keputusan pada tahap preprocessing dan pemodelan.

e. Transformasi *& Preprocessing*

Transformasi dan *preprocessing* dilakukan untuk membersihkan data sekaligus mengubah formatnya agar sesuai dengan kebutuhan algoritma. Langkah-langkah yang dilakukan adalah sebagai berikut:

1) Penanganan nilai kosong *(missing value)*: atribut numerik seperti *rating* diisi dengan nilai *median*, sedangkan atribut kategorikal seperti *genre*, *platform*, dan ESRB diisi dengan modus.
2) Penghapusan data duplikat untuk mencegah redundansi.
3) Standarisasi data numerik (*rating*): dilakukan menggunakan *StandardScaler* agar nilai berada dalam rentang yang sama.
4) *Encoding* atribut kategorikal: ESRB diubah ke bentuk numerik dengan *Label Encoding*, sedangkan *genre* dan *platform* diubah menggunakan *multi-hot encoding* karena satu *game* dapat memiliki lebih dari satu kategori.

Tahap ini penting untuk menghasilkan data yang bersih, seragam, dan dapat diproses lebih baik oleh algoritma *clustering* maupun *content-based filtering.*

f. Clustering (K-Means)

Clustering digunakan untuk mengelompokkan *game* berdasarkan

kesamaan atribut. Algoritma yang digunakan adalah *K-Means*, yang membagi dataset ke dalam sejumlah cluster sesuai nilai K yang ditentukan. Jumlah cluster optimal ditentukan menggunakan metode *Elbow* dan *Silhouette Score*, yang kemudian menghasilkan nilai K=3 sebagai jumlah *cluster* terbaik.

Hasil *clustering* menunjukkan pembagian dataset yang relatif seimbang, dengan tiga *cluster* yang masing-masing mewakili kelompok *game* dengan karakteristik tertentu. Untuk memperjelas hasil, visualisasi dilakukan menggunakan algoritma UMAP dan t-SNE, sehingga distribusi cluster dapat diamati secara visual dalam ruang dua dimensi.

g. *Content – Based Filtering* (CBF)

Metode *content-based filtering* digunakan untuk menghasilkan rekomendasi *game* berdasarkan kesamaan konten. Proses ini dilakukan dengan menghitung nilai *cosine similarity* antar *game* pada atribut *genre*, *platform*, *rating*, dan ESRB yang telah diproses.

Sebagai contoh, ketika pengguna memilih *game Grand Theft Auto V* (GTA V), sistem akan merekomendasikan *game* yang memiliki tingkat kemiripan tinggi, seperti *Skyrim, Assassin's Creed*, dan *Elden Ring*. Dengan cara ini, pengguna dapat memperoleh rekomendasi yang sesuai dengan preferensi berdasarkan konten dari *game* yang dipilih.

h. Hybrid Recommendation

Pendekatan *hybrid recommendation* diterapkan dengan menggabungkan hasil *clustering K-Means* dan *content-based filtering*. Mekanisme yang digunakan adalah membatasi pencarian rekomendasi hanya pada *game* yang berada dalam cluster yang sama, lalu menghitung

nilai kemiripan menggunakan *cosine similarity.*

Dengan cara ini, sistem tidak hanya mengandalkan kesamaan konten antar *game*, tetapi juga mempertimbangkan kesamaan dalam kelompok *cluster*. Hal ini menjadikan rekomendasi lebih relevan, bervariasi, dan sesuai dengan preferensi pengguna.

i. Implementasi Sistem Rekomendasi

Tahap terakhir penelitian adalah implementasi sistem rekomendasi ke dalam aplikasi berbasis web. Sistem dibangun menggunakan *framework* Django yang mendukung pengembangan berbasis Python. Fitur utama sistem meliputi:

1) Autentikasi pengguna (*login* dan *registrasi*).
2) Navigasi berdasarkan *genre*, *platform*, *rating*, dan ESRB.
3) Halaman *detail game* yang menampilkan informasi lengkap.
4) Fitur rekomendasi *game* menggunakan pendekatan *clustering*, *content-based filtering*, maupun *hybrid*.

Sistem yang telah selesai dibangun kemudian dideploy ke *Google Cloud Platform* (GCP) agar dapat diakses secara *real-time* melalui internet. Dengan demikian, hasil penelitian tidak hanya bersifat teoretis, tetapi juga dapat digunakan secara praktis oleh pengguna.

3.2 Tempat dan Waktu Penelitian

Dalam penelitian ini dilakukan beberapa kegiatan untuk menyelesaikan penelitian yaitu mengeidentifikasi masalah, melakukan pengumpulan data,

Preprocessing data, *clustering*, penggabungan *clustering* dan *content-based filtering*, implementasi menjadi web lalu *deployment* dan penulisan laporan. Penelitian dilakukan pada bulan Mei 2025 sampai Juli 2025, sedangkan dalam pengambilan data dilakukan pada tanggal 24 April 2025 sampai 28 April 2025.

3.3 Alat dan Bahan Penelitian

1. Spesifikasi *Hardware*

Berikut adalah spesifikasi perangkat keras *(hardware)* yang digunakan dalam penelitian ini:

a. Spesifikasi Laptop

Laptop : *Acer*
CPU : *AMD Ryzen 3 3250U with Radeon Graphics*
GPU : *AMD Radeon (TM) Vega 3 Graphics*
Sistem Operasi : *Windows 10*
Memori : 8 GB RAM
Storage : 512 GB SSD

b. Spesifikasi *Smartphone*

Smartphone : Iphone 11
CPU : *Chip A13 Bionic*
Sistem Operasi : iOS 18.3.2
Memori : 4GB RAM
Storage : 64 GB

2. Spesifikasi Software

Berikut adalah daftar perangkat lunak (*software*) yang digunakan dalam penelitian ini:

a. Google Colab

b. Microsoft Excel

c. Visual Studio Code

d. Google Cloud *Platform*

e. GitHub

3. Bahan Penelitian

Bahan penelitian yang digunakan adalah data dari API RAWG, yaitu sebuah *platform database game digital* yang menyediakan lebih dari 500.000 data *game* dari berbagai sistem dan *platform*. Data yang diambil dalam penelitian ini mencakup atribut utama, yaitu nama *game*, *genre*, *platform*, *rating*, dan kategori usia (ESRB).

Data dari RAWG dipilih karena sifatnya yang lengkap, terstruktur, serta dapat diakses secara *real-time* melalui API publik. Dengan karakteristik tersebut, data ini sesuai untuk digunakan dalam pengembangan sistem rekomendasi, karena mampu merepresentasikan keragaman *game* yang tersedia dan relevan dengan kebutuhan pengguna.

3.4 Metode Pengumpulan Data

1. Data sekunder

Data sekunder adalah data yang diperoleh bukan secara langsung oleh peneliti, melainkan bersumber dari literatur, laporan, maupun *database* yang telah tersedia sebelumnya. Data jenis ini biasanya telah dikumpulkan dan diolah oleh pihak lain, kemudian dimanfaatkan kembali untuk mendukung penelitian yang sedang dilakukan [44]. data yang diperoleh dari berbagai sumber terdokumentasi, seperti catatan, buku, jurnal, arsip, maupun laporan

historis yang telah tersedia dan tersusun sebelumnya [45]. Karena sifatnya yang sudah tersedia, data sekunder dapat membantu peneliti menghemat waktu, biaya, dan tenaga, terutama dalam penelitian yang membutuhkan data berskala besar atau bersifat longitudinal.

2. Studi Pustaka

Menurut Purba Wahyu Adi, studi pustaka adalah penelitian yang menganalisis berbagai literatur untuk membangun kerangka konseptual, memperkuat pemahaman teoretis, mengidentifikasi celah penelitian, dan mendukung dasar metodologis serta praktis penelitian [46].

Dalam penelitian ini, studi pustaka digunakan untuk mendukung pengembangan sistem rekomendasi *game* dengan meninjau teori dasar tentang sistem rekomendasi, algoritma *machine learning*, serta penelitian-penelitian terkait yang relevan. Dengan demikian, studi pustaka tidak hanya berfungsi sebagai referensi teoretis, tetapi juga sebagai pijakan praktis dalam menentukan metodologi penelitian yang tepat.

Adapun buku utama yang dijadikan rujukan dijelaskan sebagai berikut:

a. Gallatin dan Albon, dalam buku *Machine learning with Python Cookbook*, menyajikan panduan praktis mengenai penerapan algoritma *K-Means Clustering* dalam proses klasterisasi data. Buku ini memuat lebih dari 200 resep kode *Python* yang mencakup berbagai aspek penting dalam bidang *machine learning*, seperti pengolahan data, ekstraksi fitur, evaluasi model, serta penyimpanan dan *deployment* model terlatih. Penulis juga membahas metode representasi data teks seperti *Bag-of-Words* dan TF-IDF *weighting*, yang menjadi dasar teknik *Content-based Filtering*. Buku ini memberikan kerangka kerja

yang jelas dan modular dalam merancang sistem pembelajaran mesin yang efisien dan siap diterapkan dalam lingkungan produksi. Dalam konteks penelitian ini, buku tersebut dijadikan acuan utama dalam implementasi algoritma klasterisasi *K-Means* dan *Content-based Filtering* untuk membangun sistem rekomendasi *game*. Khususnya, digunakan dalam proses pemrosesan data fitur serta penyusunan *pipeline* rekomendasi secara sistematis dan terstruktur [47].

b. Atram et al., dalam buku *Data Mining*: *Concepts and Techniques,* menyajikan landasan teoritis dan praktis yang komprehensif mengenai berbagai teknik *data mining* yang esensial dalam pengembangan sistem rekomendasi. Buku ini menguraikan proses *Knowledge Discovery in Database* (KDD) secara sistematis, mulai dari tahap praproses data, eksplorasi pola, hingga evaluasi model. Salah satu fokus utama dalam buku ini adalah pembahasan mendalam terhadap teknik klasterisasi, khususnya algoritma *K-Means,* dalam konteks segmentasi data untuk mengidentifikasi kelompok-kelompok homogen dalam dataset berskala besar. Selain itu, penulis juga menjelaskan metode representasi data berbasis fitur, yang menjadi fondasi utama dari pendekatan *Content-based Filtering.* Dengan pendekatan yang sistematis dan aplikatif, buku ini memberikan panduan praktis yang bermanfaat dalam merancang dan mengimplementasikan sistem rekomendasi yang efektif, efisien, dan berbasis pada pola data yang terstruktur [48].

c. Dr. Budi Raharjo, dalam buku *Machine Learning*: Pembelajaran Mesin, memberikan pemahaman menyeluruh mengenai konsep, jenis, serta penerapan algoritma pembelajaran mesin, baik yang bersifat *supervised*

maupun *unsupervised learning*. Buku ini disusun secara sistematis dan membahas topik-topik inti seperti klasifikasi, *regresi, clustering*, pemrosesan data, serta teknik evaluasi model melalui metrik akurasi, presisi, dan validasi silang. Selain pembahasan teoritis, buku ini dilengkapi dengan pendekatan praktis berbasis bahasa pemrograman *Python* dan *library Scikit-learn*, serta disertai studi kasus seperti klasifikasi bunga iris dan prediksi data *numerik*. Hal ini menjadikannya sumber yang sangat membantu dalam pengembangan sistem pembelajaran mesin secara aplikatif. Penjelasan rinci mengenai algoritma *K-Means, K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM) memperkuat fondasi teknis bagi peneliti dalam membangun sistem rekomendasi berbasis konten yang efektif, akurat, dan dapat dievaluasi secara terukur dalam lingkungan data nyata [49].