

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring dengan perkembangan zaman, kemajuan dalam proses pengumpulan data dan teknologi penyimpanan yang cepat dan akurat memungkinkan organisasi menghimpun jumlah data yang sangat luas, sehingga penggunaan alat dan teknik analisis data secara manual tentunya tidak dapat digunakan untuk mengekstrak informasi dari data yang sangat besar.

Data mining adalah metode yang memungkinkan para penggunanya untuk mengakses data yang besar dalam waktu yang relatif cepat. Dalam menciptakan efisiensi pengumpulan data yang besar tentunya diperlukan metode baru yang dapat menjawab kebutuhan tersebut. *Data mining* merupakan sebuah teknologi yang dapat memproses data dalam volume besar yang digunakan oleh Perusahaan untuk mengubah data mentah menjadi informasi yang berguna untuk membuat suatu keputusan bisnis yang sangat penting (Mustika,2021). Sudah banyak penelitian terkait dengan *data mining* dengan berbagai metode sesuai dengan keperluan penelitian. Mulai dari *clustering*, analisis, hingga untuk keperluan prediksi berdasarkan parameter yang digunakan.

Goodreads Indonesia yang dibentuk pada tanggal 7 Juni 2007 oleh Femmy Syahrani merupakan salah satu situs yang menyediakan informasi tentang berbagai jenis buku. Situs ini ditujukan untuk para pembaca buku berbahasa Indonesia yang ingin mendiskusikan buku-buku tersebut dan juga upaya untuk mengumpulkan

buku-buku berbahasa Indonesia (*goodreads-indonesia,2023*). Banyaknya jumlah buku yang tersedia pada *Goodreads* membuat pembaca terkadang kesulitan dalam mengetahui buku yang direkomendasikan oleh pengguna lain. Dalam upaya meningkatkan peran aktif *Goodreads* Indonesia dalam dunia perbukuan, diperlukan sebuah strategi dalam hal penelusuran buku, yaitu sebuah sistem yang dapat memahami kebutuhan dan karakteristik dari pengguna *Goodreads* melalui buku yang dibaca.

K-Means merupakan algoritma klusterisasi yang memiliki ide dasar sederhana dengan cara meminimalkan *Sum of Squared Error (SSE)* antara objek-objek data dengan sejumlah *k centroid*. *K-means Clustering* merupakan salah satu algoritma yang digunakan untuk mengelompokkan data ke dalam data *cluster*. Algoritma ini dapat mengelompokkan data besar dengan waktu yang singkat. *K-Means Clustering* juga umum digunakan karena mudah dipelajari dan sederhana karena algoritma ini hanya menghitung jarak tiap data ke titik pusat *cluster* (Suyanto,2017).

Berdasarkan penelitian yang dilakukan oleh Nurul Kasanah (2023) dijelaskan bahwa pengelompokan data tingkat kemiskinan menggunakan algoritma *K-Means Clustering* berhasil menghasilkan pengelompokan data yang akurat. Hasil dari penelitian ini didapatkan bahwa dengan adanya 3 *cluster* kemiskinan dari tahun 2011 hingga tahun 2022.

Penelitian serupa juga dilakukan oleh Hesti Astuti (2021) dalam penelitian pengelompokan data pelanggan PT. Pinus Merah Abadi menggunakan algoritma *K-Means*. Hasil dari penelitian ini adalah menghasilkan tiga *cluster* pelanggan

dengan karakteristik pelanggan dari tiap-tiap *cluster*. Hasil dari *clustering* ini dapat membantu PT.Pinus Merah Abadi untuk mengambil kebijakan dan strategi dalam meningkatkan atau mempertahankan pelanggan yang sesuai dengan *cluster* pelanggan tersebut.

Dengan memanfaatkan *data mining* metode *K-Means Clustering*, setiap kategori buku yang ada di situs *Goodreads* Indonesia dapat dikelompokkan dalam jumlah *cluster* untuk perencanaan penambahan koleksi buku bacaan. Hasil yang didapatkan yaitu, kelompok buku yang akan dijadikan sebagai bahan evaluasi penelusuran oleh pengguna *goodreads* Indonesia untuk dijadikan target bacaan selanjutnya. Diharapkan dengan adanya penelitian mengenai “Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering” dapat memudahkan pembaca dalam menjelajahi jenis buku yang paling banyak diminati.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang telah penulis paparkan, maka dapat dirumuskan suatu permasalahan bagaimana mengelompokkan buku berdasarkan jumlah *cluster* yang belum ditentukan menggunakan *algoritma K-Means Clustering*.

1.3 Ruang Lingkup

Penelitian yang dilakukan memiliki batasan dan ruang lingkup sebagai berikut:

1. Algoritma *clustering* yang digunakan adalah algoritma *K-Means*

2. Sumber data didapatkan dari *Kaggle* dengan atribut judul, judul_slug, author, penerbit, isbn, bahasa, jumlah_halaman, isBestSeller, original_rating, adjusted_rating, rating_count, review_count, kategori
3. Atribut yang digunakan dalam pengelompokan data dengan *K-Means Clustering* adalah *Original Rating*, *Adjust Rating* dan *Rating Count* karena hanya berfokus pada nilai *rating*.
4. Data yang diambil berjumlah 2125 record
5. Proses analisis kelompok data dijalankan menggunakan bahasa pemrograman *python* dan library *streamlit* dengan tool *Jupyter Notebook*
6. Hasil analisis data merupakan hasil pengelompokan data dengan fitur yang memiliki kemiripan

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah

1. Menampilkan hasil berupa pengelompokan data buku yang memudahkan pengguna *goodreads* Indonesia dalam penelusuran buku berdasarkan kelompok *cluster*.
2. Mengimplementasikan algoritma *K-Means Clustering* menggunakan bahasa pemrograman *python*

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah Mengetahui penerapan *data mining* dalam pengelompokan buku menggunakan algoritma *K-Means Clustering*.

1.6 Sistematika Penulisan

Sistem penulisan skripsi dengan judul “Implementasi *Data Mining* Untuk Pengelompokan Buku Menggunakan Algoritma *K-Means Clustering*“ disusun guna memberi gambaran umum tentang penelitian yang dilakukan. Sistematika penulisan skripsi ini adalah sebagai berikut :

BAB I PENDAHULUAN

Menguraikan tentang pokok yang menjadi latar belakang permasalahan yang diambil, menguraikan rumusan masalah yang di hadapi, menentukan ruang lingkup masalah untuk membatasi aplikasi, menentukan tujuan, manfaat dan kegunaan sistem, serta sistematika penulisan

BAB II TINJAUAN PUSTAKA DAN DASAR TEORI

Membahas mengenai tinjauan pustaka yaitu mengacu penelitian-penelitian yang ada sebelumnya dengan meninjau kelebihan dan kekurangan dari penelitian tersebut sehingga dapat digunakan sebagai referensi. Sedangkan dasar teori berisi konsep dasar serta teori-teori yang berkaitan dengan topik penelitian yang digunakan untuk mendukung proses analisis permasalahan. Selain itu memberikan gambaran teknologi-teknologi yang digunakan sebagai pendukung penelitian yang akan dilakukan.

BAB III METODE PENELITIAN

Menganalisis sistem dalam aplikasi yang dibangun dimulai dari aspek-aspek yang berkaitan serta merancang sistem dimulai dari segi analisis kebutuhan, terdiri dari kebutuhan *input*, *output*, perangkat keras, dan perangkat lunak, selanjutnya

berisi tentang pemodelan-pemodelan dengan diagram-diagram, hingga berisi desain tampilan yang dapat membantu dalam proses implementasi.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Berisi kesimpulan serta menjawab permasalahan yang dihadapi sampai mengetahui keunggulan dan kekurangan dari sistem yang dirancang dengan yang sudah ada, serta hasil implementasi dalam pembuatan aplikasi ini.

BAB V PENUTUP

Menjelaskan tentang kesimpulan dan saran dari penelitian yang sudah dilakukan.