

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kunci kemajuan bagi suatu bangsa salah satunya adalah minat baca, hal ini dikarenakan membaca dapat menambah informasi dan memperluas pengetahuan sehingga dapat membentuk perilaku dan pola pikir manusia. Terbukti dari survei, peringkat teratas literasi dipegang oleh negara-negara maju seperti Amerika, Jepang, Jerman dan negara maju lainnya sudah mempunyai tradisi membaca buku. Bagi masyarakat negara tersebut, buku sudah menjadi teman yang selalu dibawa kemanapun mereka pergi [1].

Tradisi membaca buku di negara tersebut sangatlah berbeda dengan Indonesia. Menurut data dari *World's Most Literate Nations Ranked* survei dari Central Connecticut State University tentang peringkat literasi negara-negara di dunia pada tahun 2016, Indonesia berada di urutan 60 dari 61 negara yang disurvei [1]. Presentase minat baca di Indonesia masih sangat rendah dibandingkan dengan negara maju, hal ini berdasarkan data dari *United Nations Educational Scientific, and Cultural Organization (UNESCO)* minat baca masyarakat Indonesia hanya 0,001% yang artinya hanya ada 1 orang yang memiliki minat baca tinggi dari 1000 orang [2]. Data hasil dari studi IEA (*International Association for the Evaluation of Educational Achievement*) juga menyebutkan bahwa minat baca terendah dipegang oleh Indonesia dengan skor 51.7, berada dibawah Filipina 52.6, Thailand 65.1, Singapura 74.0 dan Hongkong 75.5 [3]. Berdasarkan data tersebut, diperlukan perhatian khusus untuk meningkatkan minat baca masyarakat di Indonesia.

Berdasarkan data tersebut penulis ingin membangun sebuah sistem rekomendasi untuk membantu mendorong meningkatkan minat baca. Mendapatkan buku yang tepat akan menambah keinginan dalam membaca lebih besar. Sistem rekomendasi (RS) telah muncul sebagai sistem yang populer di era big data, RS telah membantu pengguna menemukan informasi yang relevan. Ada beberapa teknik yang digunakan untuk membangun sebuah sistem rekomendasi, salah satunya adalah *collaborative filtering*.

*Collaborative filtering* memanfaatkan informasi yang didapatkan dari *user* sebelumnya yang serupa untuk merekomendasikan item kepada *user* baru, oleh karena itu teknik ini termasuk ke dalam bidang kecerdasan kolektif (*collective intelligence*). *Collaborative filtering* dapat dibangun dengan salah satu dari dua metode, yaitu *memory-based collaborative filtering* dan *model-based collaborative filtering* [4].

*Memory-based collaborative filtering* terbagi menjadi 2 pendekatan, yaitu *user-based collaborative filtering* (UB-CF) dan *item-based collaborative filtering* (IB-CF) [4]. Dalam penelitian ini, penulis menggunakan pendekatan UB-CF karena di penelitian sebelumnya menunjukkan bahwa UB-CF lebih baik di berbagai kondisi dataset yang digunakan dalam penelitian tersebut, dalam melakukan improve UB-CF juga lebih cepat seiring ditingkatkannya data training [5]. UB-CF memanfaatkan kedekatan antar *user* untuk merekomendasikan sesuatu kepada *user* lain yang mempunyai minat yang mirip. Untuk mencari kedekatan antar *user*, penulis menggunakan algoritma *k-nearest neighbor* (k-NN). K-NN adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance

yang baru diklasifikasikan berdasarkan mayoritas dari k-NN. K-NN bekerja dengan berdasarkan pada jarak terpendek dari sample uji ke sample latih untuk menentukan k-NN nya [6]. Serta untuk mengukur jarak kedekatan (*similarity*) antar *user* menggunakan metode *euclidean distance*.

Sistem rekomendasi yang baik adalah sistem yang dapat memberikan rekomendasi buku yang sesuai dengan minat pembaca. Dalam penelitian ini penulis membandingkan sistem rekomendasi yang menggunakan dataset sebelum dan sesudah dilakukan pembersihan data agar mendapatkan sistem rekomendasi yang baik. Hasil dari UB-CF adalah sebuah prediksi, maka dari itu diperlukan sebuah evaluasi untuk mengukur performa yang dihasilkan sistem rekomendasi yang telah dibangun. Dalam penelitian ini, penulis menggunakan beberapa metode evaluasi yaitu *mean absolute error* (MAE), *root mean squared error* (RMSE) dan *mean squared error* (MSE). Tiga metode tersebut dipilih untuk mengetahui nilai error yang dihasilkan dari sistem rekomendasi yang telah dibangun.

## 1.2 Rumusan Masalah

Dari latar belakang yang telah disampaikan, didapatkan beberapa rumusan masalah sebagai berikut :

1. Bagaimana langkah-langkah membangun sistem rekomendasi buku berbasis *user-based collaborative filtering* dan menggunakan algoritma *k-nearest neighbor* dengan metode *euclidean distance*?
2. Bagaimana hasil evaluasi MAE, MSE dan RMSE pada sistem rekomendasi yang telah dibangun?
3. Bagaimana hasil perbandingan akurasi dari sistem rekomendasi yang

dibangun menggunakan dataset sebelum dan sesudah dilakukan pembersihan data (*data cleaning*)?

### 1.3 Batasan Masalah

Beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Sistem rekomendasi dalam penelitian ini menggunakan metode pendekatan *user-based collaborative filtering* dengan algoritma *k-nearest neighbor*.
2. Uji coba dan evaluasi nilai K dalam sistem rekomendasi yang dibangun pada penelitian ini adalah 1, 5, 10, 15, 20, 25.
3. Dalam mencari jarak kedekatan (*similarity*) antar *user* menggunakan metode *euclidean distance*.
4. Pemrosesan sistem rekomendasi dalam penelitian ini menggunakan Google Colaboratory.
5. Menggunakan dataset *rating* buku yang diperoleh dari *website* ([www.kaggle.com](http://www.kaggle.com)).
6. Objek buku yang diolah adalah sebanyak 271360 judul buku.
7. Data *rating* yang diolah adalah sebanyak 1149780 dari 105283 *user*.

### 1.4 Maksud dan Tujuan Penelitian

Adapun maksud dan tujuan dari penelitian ini adalah:

1. Mengetahui langkah-langkah membangun sistem rekomendasi berbasis *user-based collaborative filtering* dan menggunakan algoritma *k-nearest neighbor* dengan metode *euclidean distance*.

2. Mengetahui hasil evaluasi MAE, MSE dan RMSE pada sistem rekomendasi yang telah dibangun.
3. Mengetahui hasil perbandingan akurasi dari sistem rekomendasi yang dibangun menggunakan dataset sebelum dan sesudah dilakukan pembersihan data (*data cleaning*).

### **1.5 Manfaat Penelitian**

Memberikan pengetahuan kepada pembaca tentang hasil perbandingan sistem rekomendasi yang dibangun menggunakan dataset sebelum dan sesudah dilakukan pembersihan data dengan pendekatan user-based collaborative filtering dan menggunakan algoritma k-nearest neighbor dengan metode euclidean distance, serta mengetahui nilai K terbaik dari perbandingan sistem rekomendasi tersebut.

### **1.6 Metode Penelitian**

Untuk mendapatkan data yang relevan, benar dan terarah berdasarkan permasalahan yang ada, perlu adanya macam-macam metode untuk mencapai sebuah tujuan dalam penelitian ini. Macam-macam metode yang digunakan dalam penelitian ini adalah sebagai berikut :

#### **1.6.1 Metode Studi Literatur**

Penelitian sistem rekomendasi menggunakan *collaborative filtering* dengan algoritma *k-nearest neighbor* sudah ada sebelumnya, sehingga penulis mengumpulkan data penelitian sebelumnya dari buku buku, jurnal-jurnal nasional maupun internasional dan internet yang berkaitan dengan permasalahan yang diangkat dalam penelitian ini.

### 1.6.2 Metode Data Mining

Forecasting merupakan metode data mining yang digunakan dalam penelitian ini, dimana proses-prosesnya adalah sebagai berikut :

#### 1. Pengumpulan data

Data yang digunakan dalam penelitian ini adalah data yang diambil dari situs Kaggle ([www.kaggle.com](http://www.kaggle.com)) yaitu Book-Crossing: User review ratings, data tersebut bersifat terbuka. Terdapat tiga file pada dataset tersebut yakni *bx books*, *bx-users* dan *bx-book-ratings*.

#### 2. Pemilihan Data

Tahap ini penulis melakukan pemilihan dataset yang telah dikumpulkan. Dataset yang telah terpilih, nantinya akan digunakan untuk menjadi data latih dan data uji sistem rekomendasi.

#### 3. Pembersihan Data

Tahap ini yang dilakukan penulis adalah pembersihan data terhadap dataset yang sudah terpilih. Proses pembersihan data pada penelitian ini adalah menghapus data-data yang tidak diperlukan sehingga data tersebut dapat dinyatakan ideal untuk dilatih.

#### 4. Pengelohan Data

Tahap ini yang dilakukan oleh penulis adalah mengolah dataset *rating* buku yang diberikan oleh *user* sebelumnya yang nantinya akan digunakan untuk merekomendasikan buku kepada *user* baru. Tahap pengolahan ini menggunakan pendekatan *user-based collaborative filtering* dengan menggunakan algoritma *k-nearest neighbor*.

### 1.6.3 Metode Pengujian

Tahap ini untuk mengetahui performa dari sistem rekomendasi yang telah dibangun dibutuhkannya pengujian. Tahap pengujian ini penulis menggunakan metode *mean absolute error* (MAE), *mean squared error* (MSE) dan *root mean squared error* (RMSE). MAE menghitung rata-rata nilai absolut dari selisih antara nilai aktual dengan nilai prediksi. MSE menghitung rata-rata selisih kuadrat antara nilai prediksi dengan nilai aktual. RMSE menghitung akar kuadrat dari selisih nilai rata-rata antara nilai aktual dengan nilai prediksi.

### 1.7 Sistematika Penulisan

Berikut sistematika penulisan yang bertujuan untuk memudahkan pembaca dalam memahami laporan skripsi ini :

#### BAB I PENDAHULUAN

Bab ini terdiri dari tujuh sub bab yaitu latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan untuk membantu memudahkan pembaca dalam memahami laporan skripsi ini.

#### BAB II LANDASAN TEORI

Bab landasan teori ini terdiri dari tinjauan pustaka yang diperoleh dari buku-buku, jurnal-jurnal ilmiah baik nasional dan internasional dan internet yang digunakan untuk

memperkuat dasar-dasar teori dalam pembuatan sistem rekomendasi.

### BAB III METODE PENELITIAN

Bab metode penelitian berisi mengenai tahapan langkah-langkah dan contoh perhitungan manual dalam pembuatan sistem rekomendasi dalam penelitian ini.

### BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang hasil dari tahapan penelitian yang dilakukan secara menyeluruh termasuk hasil dari evaluasi algoritma.

### BAB V PENUTUP

Bab penutup berisi mengenai kesimpulan yang berdasarkan hasil dari penelitian yang dilakukan dan memberikan saran-saran untuk perbaikan dan pengembangan yang lebih lanjut agar dapat tercapai hasil yang lebih baik.