

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Banyaknya persebaran data dan informasi yang tersedia di zaman teknologi saat ini semakin meningkat jumlahnya dari waktu ke waktu. Data dan informasi yang banyak dan tersebar ini lah mengakibatkan terjadinya *information overload* apabila tidak diatasi. Sistem rekomendasi hadir untuk menjadi salah satu alat dalam mengatasi sekaligus memanfaatkan hal tersebut. Manfaat dan kegunaan sistem rekomendasi tidak bisa dianggap remeh, berbagai bidang telah menerapkan sistem ini seperti media, kesehatan, transportasi, hingga *e-commerce*[1].

Sistem rekomendasi merupakan teknologi yang ampuh dalam menyerap maupun memilih data dan informasi menjadi suatu nilai yang berharga baik bagi perusahaan maupun pengguna. Pengguna menjadi terbantu dalam mencari barang yang diinginkan, sehingga perusahaan mampu meningkatkan penjualan produk. Salah satu teknik yang digunakan dalam sistem rekomendasi yaitu *Collaborative Filtering* yang mana dibagi menjadi 2 jenis, yaitu berdasarkan barang atau *item-based* dan berdasarkan pengguna atau *user-based*[2].

Teknik *Collaborative Filtering* menggunakan Algoritma KNN yang secara umum sering digunakan dalam membuat prediksi. Prediksi yang digunakan berdasarkan preferensi pengguna/item dengan mengidentifikasi rating dan preferensi pengguna/item yang lain. KNN menghitung jarak antara tetangga terdekat (*similarity*) menggunakan metode pengukuran jarak seperti *Pearson*, *Cosine*, *Euclidian Distance*, *Mean Squared Difference* dan masih banyak lagi[3].

Dalam sistem rekomendasi, terdapat library *scikit* yang bernama *Surprise*. Library ini berguna dalam membangun dan menganalisis sistem rekomendasi yang berhubungan dengan rating data eksplisit[4]. Dalam penelitian ini, penulis melakukan pengujian menggunakan metode bawaan library *Surprise* yakni *MSD* dan *Pearson* serta *RMSE (Root Mean Squared Error)* untuk pengukuran akurasi pada model.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka dapat diambil rumusan masalah sebagai berikut :

1. Bagaimana langkah dalam pengujian performa algoritma *similarity Mean Squared Difference* dan *Pearson Correlation Coefficient* berbasis *Item Collaborative Filtering* menggunakan Algoritma *K-Nearest Neighbor* (KNN) yang terdapat pada *library Surprise* ?
2. Berdasarkan nilai  $k$  , berapakah akurasi dari hasil evaluasi RMSE pada kedua model ?
3. Algoritma *Similarity* mana yang lebih baik antara *Mean Squared Difference* dengan *Pearson Correlation Coefficient* dalam pengujian menggunakan Algoritma *K-Nearest Neighbor* (KNN) dengan teknik evaluasi *Root Mean Squared Error* (RMSE) yang terdapat pada *library Surprise* ?

## 1.3 Batasan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka dapat diambil rumusan masalah sebagai berikut :

1. Perbandingan metode pada penelitian ini berbasis item dengan Algoritma *K-Nearest Neighbour*.
2. Penelitian ini menggunakan *Pearson Correlation Coefficient* dan *Mean Squared Difference (MSD)* sebagai pengujian perbandingan nilai kedekatan.
3. Nilai  $K$  yang digunakan berjumlah 1 sampai 10.
4. Perbandingan performa metode pada penelitian ini dibuat menggunakan visual studio code dan bahasa pemrograman *Python* serta dengan bantuan *library Surprise*.
5. Dataset yang digunakan untuk pengujian bernama Amazon Review Dataset 2018, diperoleh dari website yang dipublikasikan oleh Jianmo Ni dengan

link [https://cseweb.ucsd.edu/~jmcauley/datasets/amazon\\_v2/](https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/). Penelitian ini hanya berfokus menggunakan kategori *software* pada dataset tersebut.

6. Dataset yang digunakan sebesar 25000 baris data.
7. *Root Mean Square Error (RMSE)* digunakan pada penelitian ini untuk mengukur tingkat akurasi model.

#### 1.4 Tujuan Penelitian

Tujuan adanya penelitian ini untuk membandingkan hasil pengukuran akurasi pada sistem rekomendasi menggunakan teknik KNN dengan tingkat kesamaan *Pearson Correlation Coefficient* dan *Mean Squared Difference*.

#### 1.5 Manfaat Penelitian

Terdapat berbagai manfaat yang diberikan dari adanya penelitian ini adalah sebagai berikut :

1. Mencari nilai kesamaan atau *similarity* yang terbaik antara *Mean Squared Difference* dan *Pearson Correlation Coefficient*.
2. Melakukan langkah-langkah dalam menguji akurasi metode pada sistem rekomendasi berbasis *item collaborative filtering* yang menggunakan Algoritma *K-Nearest Neighbour (KNN)* berdasarkan nilai kesamaan *MSD* dan *Pearson*.
3. Menjelaskan hasil evaluasi menggunakan *RMSE* dengan nilai *k* yang ditetapkan.

#### 1.6 Sistematika Penulisan

Penulisan pada penelitian ini terdapat sistematika penulisan yang berguna untuk menjelaskan atau mendiskripsikan secara singkat isi dari tiap-tiap bab. Berikut merupakan susunan dari sistematika penulisan :

**BAB I** PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

**BAB II** TINJAUAN PUSTAKA

Berisi studi literatur dan dasar-dasar teori yang digunakan penulis sebagai dasar dan landasan dalam penelitian ini.

**BAB III** METODE PENELITIAN

Pada bab ini berisi alat/instrumen penelitian, penjelasan dari metode penelitian yang diterapkan serta alur dari implementasi penelitian yang akan dilakukan.

**BAB IV** HASIL DAN PEMBAHASAN

Berisi tentang bagaimana cara peneliti melakukan implementasi secara teoritis dan matematis, sesuai dengan alur yang telah ditetapkan.

**BAB V** PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan, serta saran pengembangan yang bisa dilakukan pada penelitian selanjutnya.