

## **BAB 1**

### **PENDAHULUAN**

Bab ini menjabarkan latar belakang pemilihan topik Tugas Akhir, tujuan yang akan dicapai dalam Tugas Akhir, ruang lingkup penelitian, tahapan yang dilakukan dalam penelitian, dan sistematika penyajian laporan Tugas Akhir.

#### **1.1 Latar Belakang**

*Recommendation System* adalah sistem yang dirancang untuk merekomendasikan *item* yang paling relevan (misalnya produk, artikel, acara, dan film) kepada *user* (misalnya pelanggan, pengunjung, pengguna aplikasi, dan pembaca). Secara umum, pendekatan yang digunakan sebagai desain *recommendation system* dapat dikelompokkan menjadi 3, antara lain *collaborative filtering*, *content-based filtering*, dan *hybrid filtering* [1]. Pada *collaborative filtering*, rekomendasi diberikan kepada seorang *user* berdasarkan *item* yang pernah dipilih oleh *user* lain [2]. Sementara itu, pada *content-based filtering*, rekomendasi diberikan dengan menyesuaikan properti *item* preferensi seorang *user* dengan properti *item* preferensi *user* lain. Contoh properti yaitu deskripsi dan judul *item* [3]. Pada kelompok *hybrid filtering*, rekomendasi diberikan dengan memanfaatkan gabungan beberapa metode rekomendasi [4].

Pada masa awal perkembangan sistem rekomendasi, sistem rekomendasi didasarkan pada metode seperti *clustering* [5], *nearest neighbor* [6] dan *matrix factorization* [7]. Namun dalam beberapa tahun terakhir, *deep learning* [8] menunjukkan hasil yang sangat baik di berbagai domain, mulai dari pengenalan gambar hingga pemrosesan bahasa alami, demikian halnya dengan sistem rekomendasi. Faktanya sistem rekomendasi terbaik seperti yang ada di Youtube dan Amazon [9] didukung oleh *deep learning*. Pada dasarnya model *deep learning* memiliki struktur yang sangat kompleks dibandingkan dengan model *machine learning* pada umumnya [10]. Struktur model *deep learning* dikatakan kompleks karena model tersebut tersusun dari banyak parameter *weight* dan *bias* yang dipengaruhi oleh beberapa *hyperparameter* seperti *hidden layer*, lebar *hidden layer*, jumlah *node*, jumlah *filter*, ukuran *filter* dan fungsi aktivasi. Berdasarkan analisis

dan studi literatur pada penelitian [4], [6], [7], [11], bahwa aturan asosiasi sudah sesuai untuk penelitian ini dan aturan asosiasi tersebut dapat mencapai performa yang baik. Ide menambang aturan asosiasi (analisis aturan asosiasi) dan menggunakan aturan asosiasi yang didapatkan untuk sistem rekomendasi berasal dari pengamatan bahwa aturan seperti "90% pengguna yang menyukai restoran A dan restoran B juga menyukai restoran C, 30% dari semua pengguna menyukai semuanya" dapat digunakan dalam pemberian rekomendasi. Aturan tersebut disebut sebagai asosiasi restoran. Asosiasi restoran mewakili hubungan antar restoran.

Berdasarkan penelitian yang dilakukan oleh [12], Apriori dinyatakan memiliki waktu eksekusi yang lebih lama daripada *Eclat* dan *Frequent Pattern Growth*. Apriori adalah algoritma penambangan *frequent itemset* yang mudah dimengerti. Karena itu, *Apriori* adalah algoritma yang populer untuk menambang *frequent itemset*. Namun, *Apriori* memiliki masalah skalabilitas dan menghabiskan memori yang tersedia jauh lebih cepat daripada *Eclat* dan *FP-Growth*, karena itu *Apriori* tidak disarankan untuk digunakan pada kumpulan data yang besar. Meskipun *Eclat* dan *FP-Growth* memiliki performa yang hampir sama, dalam penelitian [12] *FP-Growth* menunjukkan performa yang sedikit lebih baik daripada *Eclat*.

Penelitian terdahulu yang membahas terkait sistem rekomendasi adalah penelitian Gandhe yang mengajukan *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF), *Linear SVM*, dan *Logistic Regression* untuk membangun sistem rekomendasi. Hasil evaluasi menunjukkan akurasi terbaik didapatkan dengan menggunakan *Linear SVM* [14]. Penelitian serupa dilakukan oleh Varatharajan yang mengajukan *Matrix Factorization* (MF), Funk MF, *Singular Value Decomposition* SVD++, dan *Hybrid*. Hasil evaluasi menunjukkan bahwa sistem rekomendasi berhasil memberikan rekomendasi yang sesuai untuk pengguna [14]. Penelitian lainnya yang berjudul *Mining association rule based on the disease population for recommendation of medicine needs* dalam penelitian ini, peneliti menganalisis resep pasien untuk mengidentifikasi hubungan antara penyakit dan obat yang digunakan oleh dokter dalam mengobati penyakit pasien. Kerangka

analisis meliputi: (1) pengumpulan data resep pasien, (2) penerapan k-means clustering untuk mengklasifikasikan 10 penyakit teratas, (3) penerapan algoritma Apriori untuk menemukan aturan asosiasi berdasarkan *support*, *confidence* dan *lift value*. Hasil pengujian ini membuktikan bahwa nilai dari *support*, *confidence* dan *lift* dapat dijadikan sebagai rekomendasi pemilihan obat yang tepat [11]. Penelitian dengan berjudul *Multidimensional Association Rule Optimization using Genetic Algorithm for hotspot occurrence Classification* menggunakan *genetic Algorithm* untuk mengoptimalkan pembentukan aturan asosiasi yang dihasilkan dari algoritma apriori. Hasil yang diperoleh dari penelitian ini adalah jumlah aturan dapat dikurangi. Jumlah aturan yang dihasilkan yaitu sebanyak 121 aturan pada generasi ke-300 hingga mencapai 108 aturan pada generasi ke-50 [15].

*Genetic algorithm* (GA) adalah suatu algoritma pencarian yang berbasis pada mekanisme seleksi alam dan genetika. *Genetic algorithm* merupakan salah satu algoritma yang sangat tepat digunakan dalam menyelesaikan masalah optimasi kompleks seperti banyak variabel (kontinu, diskrit atau campuran keduanya). Sifat *genetic algorithm* adalah mencari kemungkinan dari calon solusi untuk mendapatkan solusi yang optimal dalam penyelesaian masalah. Ruang cakupan dari semua solusi yang layak, yaitu berbagai objek diantara solusi yang sesuai, yang dinamakan ruang pencarian. Tiap titik di dalam ruang pencarian mempresentasikan satu solusi yang layak. Tiap solusi yang layak dapat ditandai dengan nilai fitnessnya. Solusi yang dicari dalam *genetic algorithm* adalah titik (satu atau lebih) diantara solusi yang layak dalam ruang pencarian.

Oleh karena itu, penulis mengajukan sebuah model alternatif sistem rekomendasi yang dapat menyelesaikan masalah tersebut. Model yang diajukan oleh penulis adalah kombinasi dari algoritma *FP-Growth* dan *Genetic algorithm*. Alasan penelitian ini menggabungkan *FP-Growth* dan *Genetic algorithm* adalah karena *FP-Growth* menggunakan algoritma *depth first search* (DFS) untuk mencari aturan asosiasi dalam suatu set data transaksi. Pada dasarnya, tidak ada jaminan DFS akan selalu menemukan solusi yang paling baik atau dalam kata lain DFS tidak optimal. Disamping itu, jumlah aturan asosiasi yang ditambah dengan menggunakan *FP-Growth* akan semakin bertambah seiring dengan semakin besarnya ukuran set data.

Oleh karena itu, penulis mengajukan *Genetic Algorithm* untuk mengoptimasi kumpulan aturan rekomendasi dengan cara mempertahankan / memilih aturan asosiasi yang paling sesuai dari set data transaksi.

## 1.2 Pertanyaan Penelitian

Berdasarkan latar belakang, maka pertanyaan dari penelitian ini adalah bagaimana mendapatkan aturan rekomendasi menggunakan algoritma *Frequent Pattern Growth* dan mengoptimasi kumpulan aturan rekomendasi dengan *Genetic algorithm* untuk membangun model sistem rekomendasi.

## 1.3 Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengoptimasi kumpulan aturan rekomendasi dengan cara mempertahankan atau memilih aturan rekomendasi yang paling sesuai dan memfilter aturan rekomendasi yang kurang sesuai.

## 1.4 Ruang Lingkup

Ruang lingkup penelitian adalah sebagai berikut.

- a. Data yang digunakan dalam penelitian ini adalah *Zomato Restaurants Hyderabad* oleh Chirag\_ISB yang diperoleh dari Kaggle. Kedua dataset tersebut digunakan karena terdapat unsur atribut yang memenuhi dalam pelaksanaan penelitian ini seperti nama restoran, *rating*, dan *reviewer*.
- b. Domain sistem rekomendasi adalah sistem rekomendasi restoran.

## 1.5 Sistematika Penulisan

Dokumen tugas akhir ini terdiri dari tujuh (7) bab sebagai berikut.

- 1) Bab 1 Pendahuluan, menjabarkan latar belakang, pertanyaan penelitian, tujuan penelitian, ruang lingkup penelitian, serta sistematika penulisan dokumen.
- 2) Bab 2 Landasan Teori, menjabarkan landasan teori yang digunakan untuk mendukung penyelesaian penelitian, antara lain teori terkait metode, teknik, proses, prosedur, dan perangkat (*tools*) yang terkait dengan tema penelitian. Di samping itu, pada bab ini juga terdapat kajian mengenai penelitian sebelumnya yang terkait dengan tema penelitian.

- 3) Bab 3 Analisis, menjabarkan analisis tahapan penelitian, data, teknik pra pemrosesan data, metode analisis asosiasi, metode optimasi, dan matrik evaluasi aturan asosiasi.
- 4) Bab 4 Desain, menjabarkan arsitektur sistem, desain pra pemrosesan data, desain analisis asosiasi dan optimasi, desain evaluasi aturan asosiasi, serta desain eksperimen penelitian.
- 5) Bab 5 Implementasi, menjabarkan implementasi yang telah dilakukan berdasarkan desain.
- 6) Bab 6 Hasil dan Pembahasan, menjabarkan hasil yang diperoleh dari implementasi dan pembahasannya.
- 7) Bab 7 Kesimpulan dan Saran, menjabarkan kesimpulan berdasarkan hasil penelitian dan saran penelitian untuk penelitian selanjutnya.

