

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, jumlah informasi yang tersedia di internet berkembang dengan sangat pesat. Untuk membantu pengguna dalam menemukan konten yang relevan, di tengah lonjakan informasi tersebut berbagai platform telah menerapkan sistem rekomendasi. Secara umum, sistem rekomendasi bertujuan untuk menyarankan item kepada pengguna berdasarkan sejumlah faktor, seperti kemiripan antar item, *rating*, popularitas, dan preferensi pengguna.

Taksonomi sistem rekomendasi dapat dibagi menjadi dua kategori utama, yaitu *personalized recommender system* dan *non-personalized recommender system* [1]. Pada *non-personalized recommender system* saat sistem memberikan rekomendasi, pendekatan ini berfokus pada tren umum dan memperlakukan semua penggunanya sama tanpa mempertimbangkan preferensi mereka. Contoh implementasi dari *non-personalized recommender system* adalah *Top Charts*, *Popular Lists*, *Best Seller*, *Trending* dan *Recently Added*. Sebaliknya *personalized recommender system* akan mempertimbangkan preferensi pengguna. Rekomendasi yang diberikan oleh sistem berdasarkan pada riwayat interaksi pengguna dengan item dan atau kemiripan preferensi antar pengguna. Pendekatan ini dinilai lebih efektif dalam meningkatkan relevansi dan kepuasan pengguna.

Pada *personalized recommender system* setidaknya terdapat lima pendekatan dalam memenuhi tujuan dari sistem rekomendasi. Pendekatan tersebut di antaranya *content-based filtering*, *collaborative filtering*, *knowledge based*, *demographic* dan *hybrid recommender system* [2]. Masing-masing pendekatan tersebut memiliki karakteristik yang berbeda dalam memberikan rekomendasi item yang sesuai dengan preferensi pengguna. Dalam *content-based filtering*, sistem akan memberikan rekomendasi yang mirip dengan item yang pengguna sukai sebelumnya, sedangkan dalam *collaborative filtering* sistem akan merekomendasikan item kepada pengguna aktif berdasarkan kemiripan preferensi dengan pengguna lain [3]. Adapun sistem rekomendasi dengan pendekatan *knowledge based* akan memberikan rekomendasi item berdasarkan kesamaan antara kebutuhan atau spesifikasi kebutuhan pengguna dengan deskripsi item.

Dalam *demographic*, digunakan data demografis pengguna sebagai dasar dalam merekomendasikan item [2]. Berbeda dengan metode lain, pendekatan *hybrid recommender system* dalam implementasinya akan menggabungkan beberapa metode dalam memberikan rekomendasi tujuannya untuk menutupi kelemahan dari satu metode menggunakan metode yang lain.

Salah satu tantangan utama dalam sistem rekomendasi berbasis *collaborative filtering* adalah *cold start problem*, yaitu kondisi ketika sistem kesulitan atau bahkan tidak dapat memberikan rekomendasi kepada pengguna baru (*new user cold start problem*) dan sistem tidak dapat menyertakan item baru pada daftar rekomendasinya (*new item cold start problem*) [4]. Permasalahan ini terjadi karena tidak tersedianya informasi historis yang cukup untuk membangun preferensi pengguna atau karakteristik item. *Cold start problem* secara khusus menjadi tantangan dalam implementasi *collaborative filtering*, yang mengandalkan data interaksi pengguna.

Cold start problem merupakan masalah signifikan dalam sistem rekomendasi karena menghambat kemampuan sistem dalam memberikan rekomendasi yang relevan kepada pengguna baru. Tanpa solusi untuk mengatasi masalah ini, pengguna baru sering kali menerima rekomendasi yang tidak sesuai dengan preferensi mereka, yang berisiko mengurangi tingkat relevansi antara rekomendasi item dengan item yang dibutuhkan/diinginkan pengguna. Oleh karena itu, mengatasi *cold start problem* adalah langkah yang sangat penting dalam pengembangan sistem rekomendasi yang efektif dan berkelanjutan [5].

Untuk mengatasi *new user cold start problem*, pembentukan profil awal pengguna menjadi langkah penting. Terdapat beberapa pendekatan yang dapat dilakukan dalam membentuk profil awal pengguna baru di antaranya yaitu meminta pengguna baru untuk menyebutkan preferensi mereka secara eksplisit, atau meminta pengguna baru untuk memberikan nilai (*rating*) kepada beberapa item di awal, selain itu mengumpulkan data demografis dari pengguna baru dapat dilakukan untuk mengatasi *new user cold start problem* [6]. Sedangkan untuk mengatasi *new item cold start problem* diperlukan sebuah pendekatan lain yang secara langsung membandingkan kesamaan antara item baru dengan item yang merupakan preferensi pengguna.

Beberapa penelitian telah dilakukan untuk mengatasi *cold start problem*. Penelitian yang dilakukan oleh Dyah Wardani melakukan pendekatan demografis sekaligus mengimplementasikan *Artificial Neural Network* (ANN) dalam mengatasi *new user cold start problem*. Penelitian tersebut memanfaatkan data demografis dalam membentuk preferensi pengguna [4]. Selain itu, penelitian yang dilakukan oleh J. Herce-Zelaya, C. Porcel, J. Bernabe-Moreno, A. Tejada-Lorente dan E. Herrera-Viedma melakukan pendekatan yang sama yaitu demografis dengan cara mengumpulkan informasi dari media sosial dan mengimplementasikan algoritma *random decision forest* dalam mengatasi *cold start problem* [7].

Sedangkan pada penelitian ini akan dilakukan pendekatan ekstraksi kata kunci (*keyword extraction*) dari preferensi eksplisit pengguna. Hasil dari ekstraksi kata kunci tersebut digunakan sebagai *input* pada *hybrid recommender system* yang merupakan kombinasi dari *knowledge based* dengan *collaborative filtering*. Dengan menggabungkan kedua pendekatan ini, sistem diharapkan dapat memberikan rekomendasi yang lebih relevan bagi pengguna baru dan dapat merekomendasikan item baru kepada pengguna, serta secara efektif memitigasi dampak dari *cold start problem*. Berdasarkan latar belakang yang telah dibahas sebelumnya, penelitian ini diajukan dengan judul “Mengatasi *Cold Start Problem* Pada Sistem Rekomendasi *Collaborative Filtering* Menggunakan Pendekatan *Hybrid* dan *Keyword Extraction*”.

1.2 Rumusan Masalah Penelitian

Dengan latar belakang yang telah diuraikan, pernyataan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana mengimplementasikan model ekstraksi kata kunci untuk memproses preferensi eksplisit pengguna?
2. Bagaimana mengintegrasikan hasil ekstraksi kata kunci tersebut ke dalam sistem rekomendasi berbasis *collaborative filtering* untuk mengatasi *cold start problem*?

1.3 Tujuan Penelitian

Berdasarkan pernyataan masalah yang telah disebutkan di atas, adapun tujuan dari penelitian ini adalah:

1. Mengimplementasikan model *Keyword Extraction using Bidirectional Encoder Representations from Transformers* (KeyBERT) dan *Yet Another Keyword Extractor* (YAKE) untuk mengekstraksi kata kunci yang representatif dari preferensi eksplisit pengguna.
2. Mengembangkan mekanisme integrasi antara hasil ekstraksi kata kunci dengan sistem rekomendasi berbasis *collaborative filtering* untuk membangun profil awal pengguna sekaligus memberikan rekomendasi kepada pengguna lama maupun pengguna baru.

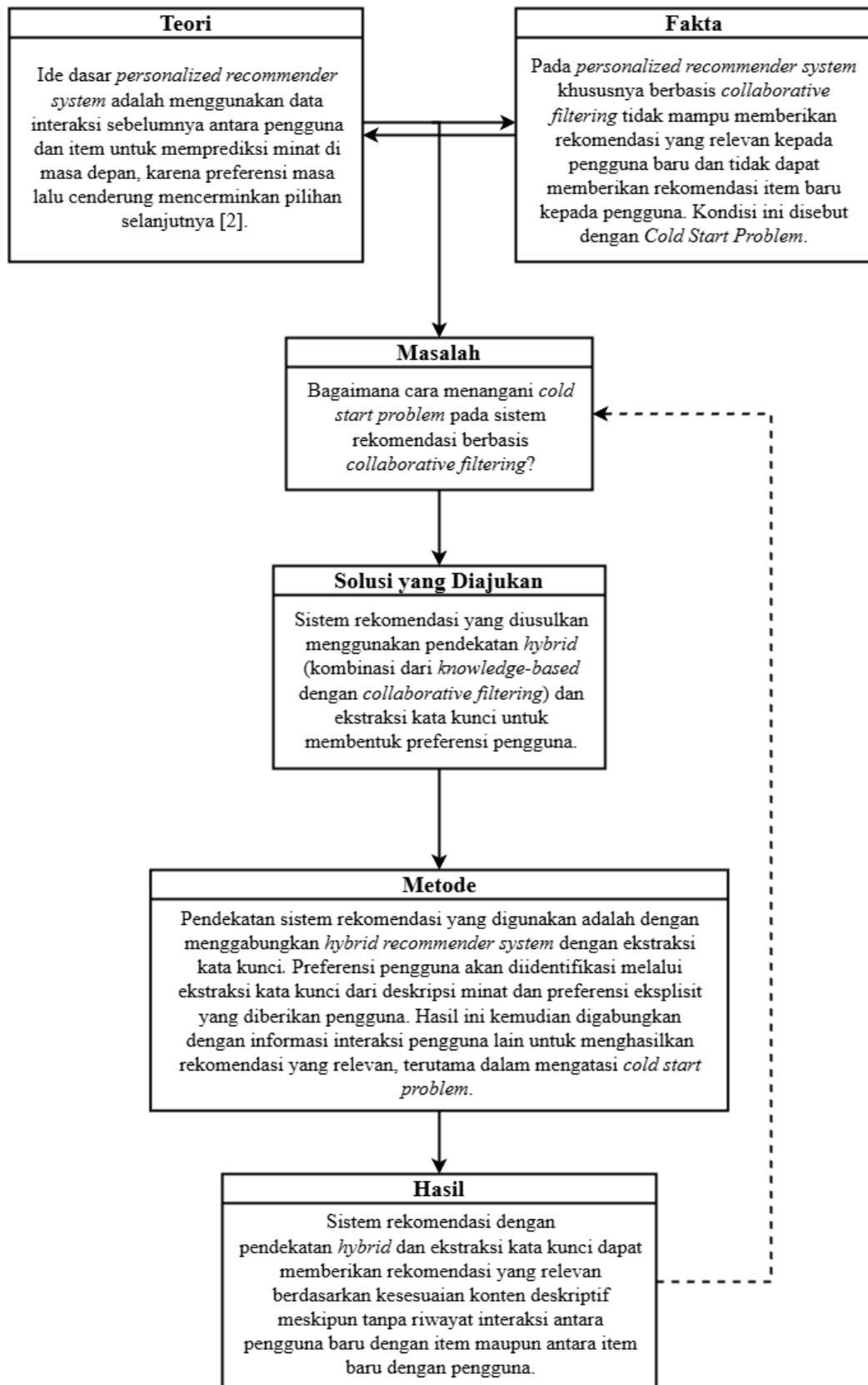
1.4 Batasan Masalah Penelitian

Untuk menjaga fokus penelitian agar tidak terlalu meluas dari pembahasan yang dimaksud, penelitian ini membatasi ruang lingkupnya. Oleh karena itu, batasan masalah dalam penelitian ini meliputi:

1. Dataset yang digunakan pada penelitian ini adalah Indonesia Tourism Destination bersumber dari penyedia dataset yang bersifat *open source* yaitu Kaggle [8]. Dataset ini terdiri data 300 pengguna, 437 objek destinasi dengan total interaksi antara pengguna dan item berjumlah 10.000 data.
2. *Hybrid recommender system* yang akan diimplementasikan pada penelitian ini adalah kombinasi dari *knowledge based* dengan *collaborative filtering*.
3. Ekstraksi kata kunci dari preferensi eksplisit pengguna dilakukan menggunakan model KeyBERT dan YAKE tanpa modifikasi arsitektur model.
4. Bahasa yang digunakan untuk *input* preferensi pengguna terbatas pada bahasa Indonesia.
5. Penelitian ini mencakup lima fase dari CRISP-DM, meliputi *business understanding*, *data understanding*, *data preparation*, *modeling* dan *evaluation*. Adapun fase *deployment* tidak termasuk dalam ruang lingkup penelitian ini.

1.5 Kerangka Pemikiran Penelitian

Kerangka pemikiran merupakan suatu model konseptual yang didalamnya berisi variabel-variabel yang berkaitan dalam sebuah penelitian. Berikut adalah kerangka pemikiran dari penelitian ini:



Gambar 1. 1 Kerangka Pemikiran

Pada Gambar di atas (Gambar 1.1) menunjukkan bahwa penelitian ini berangkat dari konsep *personalized recommender system* yang memanfaatkan data historis pengguna untuk memberikan rekomendasi yang relevan. Namun, pendekatan *collaborative filtering* memiliki keterbatasan ketika dihadapkan pada pengguna baru dan item baru yang belum memiliki riwayat interaksi, yang dikenal sebagai *cold start problem*. Masalah ini menghambat sistem dalam memberikan rekomendasi yang relevan bagi pengguna baru sekaligus menghambat sistem dalam merekomendasikan item baru. Untuk mengatasi hal tersebut, penelitian ini mengusulkan untuk menggabungkan teknik ekstraksi kata kunci dengan sistem rekomendasi berbasis *hybrid* yang merupakan kombinasi antara pendekatan *knowledge based* dengan *collaborative filtering*. Preferensi awal pengguna dibentuk menggunakan model ekstraksi kata kunci dari preferensi eksplisit yang diberikan pengguna. Hasil ekstraksi ini digunakan untuk menghasilkan preferensi item pengguna, yang kemudian diperkaya dengan data pengguna lain melalui pendekatan *collaborative filtering*. Dengan pendekatan ini, sistem diharapkan dapat memberikan rekomendasi yang relevan meskipun tanpa riwayat interaksi sebelumnya, sehingga secara efektif mengatasi *cold start problem*.

1.6 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun guna memberikan gambaran menyeluruh mengenai tahapan-tahapan yang dilakukan dalam menyelesaikan penelitian. Adapun sistematika penulisan ini terdiri dari:

BAB I PENDAHULUAN

Bab ini memuat uraian mengenai latar belakang masalah, perumusan masalah, tujuan penelitian, batasan masalah, kerangka pemikiran, serta sistematika penulisan secara keseluruhan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi kajian terhadap teori-teori dan konsep dasar yang relevan dengan topik penelitian. Seluruh teori yang disajikan mendukung proses analisis terhadap permasalahan yang diangkat.

BAB III METODOLOGI PENELITIAN

Bab ini membahas tahapan metodologis yang digunakan selama proses penelitian, mulai dari tahap analisis hingga perencanaan implementasi sistem.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil penelitian yang diperoleh berdasarkan tahapan-tahapan yang telah dilakukan. Selanjutnya, hasil tersebut dianalisis dan dibahas untuk menjawab rumusan masalah yang telah dirumuskan sebelumnya.

BAB V SIMPULAN DAN SARAN

Bab ini menyajikan kesimpulan dari penelitian yang telah dilakukan, serta memberikan kritik dan saran sebagai refleksi dan bahan pertimbangan untuk penelitian selanjutnya.

