



**CHATBOT BERBASIS KECERDASAN BUATAN DENGAN SISTEM
REKOMENDASI HYBRID UNTUK REKOMENDASI
DESTINASI WISATA**

SKRIPSI

Oleh:

ROBERT ANANDA ARITONANG

20211000038

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS BUDDHI DHARMA
TANGERANG**

2025



**CHATBOT BERBASIS KECERDASAN BUATAN DENGAN SISTEM
REKOMENDASI HYBRID UNTUK REKOMENDASI
DESTINASI WISATA**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana pada
Program Studi Teknik Informatika Fakultas Sains dan Teknologi**

**Universitas Buddhi Dharma
Jenjang Pendidikan Strata 1**

Oleh:

ROBERT ANANDA ARITONANG

20211000038

FAKULTAS SAINS DAN TEKNOLOGI

**UNIVERSITAS BUDDHI DHARMA
TANGERANG**

2025

LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini.

NIM	: 20211000038
Nama	: Robert Ananda Aritonang
Jenjang Studi	: Strata 1
Program Studi	: Teknik Informatika
Peminatan	: Data Science

Dengan ini menyetujui untuk memberikan ijin kepada pihak Universitas Buddhi Dharma, Hak Bebas Royalti Non – Eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah kami yang berjudul: “CHATBOT BERBASIS KECERDASAN BUATAN DENGAN SISTEM REKOMENDASI HYBRID UNTUK REKOMENDASI DESTINASI WISATA”, beserta alat yang diperlukan (apabila ada).

Dengan Hak Bebas Royalti Non – Eksklusif ini pihak Universitas Buddhi Dharma berhak menyimpan, mengalih-media atau format-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya, dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai Robert Ananda Aritonang atau pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Universitas Buddhi Dharma, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 04-08-2025
Yang membuat pernyataan,



Robert Ananda Aritonang
20211000038

LEMBAR PENGESAHAN PEMBIMBING
CHATBOT BERBASIS KECERDASAN BUATAN DENGAN SISTEM
REKOMENDASI HYBRID UNTUK REKOMENDASI
DESTINASI WISATA

Dibuat Oleh:

NIM : 20211000038

Nama : Robert Ananda Aritonang

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian

Komprehensif

Program Studi Teknik Informatika

Data Science

Tahun Akademik 2024/2025

Disahkan oleh,

Tangerang, 04-08-2025

Pembimbing,



Aditiya Hermawan, M.Kom., M.M

NUPTK. 9538766667130223

LEMBAR PENGESAHAN SKRIPSI
CHATBOT BERBASIS KECERDASAN BUATAN DENGAN SISTEM
REKOMENDASI HYBRID UNTUK REKOMENDASI
DESTINASI WISATA

Dibuat Oleh:

NIM : 20211000038

Nama : Robert Ananda Aritonang

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian

Komprehensif

Program Studi Teknik Informatika

Data Science

Tahun Akademik 2024/2025

Disahkan oleh,

Tangerang, 04-08-2025

Dekan,

Ketua Program Studi,



Dr. Yakub, M.Kom., M.M.

NUPTK. 1836747648130172



Hartana Wijaya, M.Kom.

NUPTK. 3844759660131192

LEMBAR PENGESAHAN TIM PENGUJI

Nama : Robert Ananda Aritonang
NIM : 20211000038
Fakultas : Sains dan Teknologi
Judul Skripsi : CHATBOT BERBASIS KECERDASAN BUATAN
DENGAN SISTEM REKOMENDASI HYBRID UNTUK
REKOMENDASI DESTINASI WISATA

Dinyatakan LULUS setelah mempertahankan di depan Tim Penguji pada hari Senin, 04-08-2025

Nama penguji:

Tanda Tangan:

Ketua Sidang : **Benny Daniawan, M.Kom**

NUPTK. 8756768669130412

Penguji I : **Edy, ST., M.Kom**

NUPTK. 7560760661130203

Penguji II : **Aditiya Hermawan, M.Kom.,
M.M**

NUPTK. 9538766667130223

Mengetahui,

Dekan Fakultas Sains dan Teknologi



Dr. Yakub, M.Kom., M.M.

NUPTK. 1836747648130172

KATA PENGANTAR

Dengan mengucapkan Puji Syukur kepada Tuhan Yang Maha Esa, yang telah memberikan Rahmat dan karunia-Nya kepada penulis sehingga dapat menyusun dan menyelesaikan Skripsi ini dengan judul **“CHATBOT BERBASIS KECERDASAN BUATAN DENGAN SISTEM REKOMENDASI HYBRID UNTUK REKOMENDASI DESTINASI WISATA”**. Tujuan utama dari pembuatan Skripsi ini adalah sebagai salah satu syarat kelengkapan dalam menyelesaikan program pendidikan Strata 1 Program Studi Teknik Informatika di Universitas Buddhi Dharma. Dalam penyusunan Skripsi ini penulis banyak menerima bantuan dan dorongan baik moril maupun materiil dari berbagai pihak, maka pada kesempatan ini penulis menyampaikan RASA terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Limajatini, S.E., M.M., B.K.P, sebagai Rektor Universitas Buddhi Dharma.
2. Bapak Dr. Yakub, M.Kom., M.M, Dekan Fakultas Sains dan Teknologi.
3. Bapak Hartana Wijaya, M.Kom, sebagai Ketua Program Studi Teknik Informatika.
4. Bapak Aditiya Hermawan, S.Kom., M.Kom., M.M, sebagai pembimbing yang telah membantu dan memberikan dukungan serta harapan untuk menyelesaikan penulisan Skripsi ini
5. Orang tua dan keluarga yang selalu memberikan dukungan baik moril dan materiil.
6. Teman-teman yang selalu membantu dan memberikan semangat

Serta semua pihak yang terlalu banyak untuk disebutkan satu-persatu sehingga terwujudnya penulisan ini. Penulis menyadari bahwa penulisan Skripsi ini masih belum sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan di masa yang akan datang.

Akhir kata semoga Skripsi ini dapat berguna bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Tangerang, 04-08-2025

Penulis

ABSTRAK

Sektor pariwisata memegang peran penting dalam mendukung pertumbuhan ekonomi, pelestarian budaya, serta peningkatan kesejahteraan masyarakat lokal. Di era digital saat ini, integrasi teknologi menjadi hal krusial dalam memaksimalkan potensi sektor ini. Salah satu pendekatan inovatif yang berkembang adalah penggunaan *Artificial Intelligence (AI)* dalam bentuk *chatbot* yang dikombinasikan dengan sistem rekomendasi. *Chatbot* berbasis AI memungkinkan interaksi yang lebih natural dengan pengguna serta dapat menyajikan rekomendasi yang sesuai dengan preferensi individu melalui teknologi *Natural Language Processing (NLP)*. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi tempat wisata berbasis *chatbot* yang mampu memberikan saran destinasi secara personal berdasarkan preferensi pengguna. Sistem ini memanfaatkan kombinasi pendekatan *Collaborative Filtering (CF)* menggunakan algoritma *SVD++*, *Content-Based Filtering (CBF)* berbasis *TF-IDF* dan *cosine similarity*, serta *Knowledge-Based Recommender* sebagai mekanisme penyaring berdasarkan kategori, lokasi, dan *rating* minimum yang diinginkan pengguna. Untuk meningkatkan fleksibilitas dan akurasi rekomendasi, pendekatan *Hybrid Filtering* diterapkan melalui mekanisme *weighted hybrid*, yang menggabungkan prediksi dari CF dan CBF dengan bobot tertentu. Sistem juga menerapkan metode *switching* untuk memilih model rekomendasi terbaik berdasarkan histori interaksi pengguna. Evaluasi model menunjukkan bahwa pendekatan hybrid menghasilkan nilai RMSE sebesar 0.6041 dan MAE sebesar 0.4185, lebih baik dibandingkan model CF dan CBF secara individu. Selain itu, integrasi dengan *chatbot* menggunakan *RASA Framework* memungkinkan pengguna berinteraksi secara *real-time* melalui percakapan untuk memperoleh rekomendasi wisata yang relevan. Hasil penelitian ini menunjukkan bahwa sistem rekomendasi yang dibangun tidak hanya mampu menangani kasus *cold-start*, tetapi juga memberikan pengalaman pengguna yang lebih responsif dan kontekstual.

Kata Kunci: Sistem Rekomendasi, *Chatbot*, *Hybrid Filtering*, *Collaborative Filtering*, *Content-Based Filtering*, *RASA Framework*

ABSTRACT

The tourism sector plays a vital role in supporting economic growth, cultural preservation, and the improvement of local community welfare. In the current digital era, the integration of technology has become crucial in maximizing the potential of this sector. One emerging innovative approach is the use of Artificial Intelligence (AI) in the form of chatbots combined with recommendation systems. AI-powered chatbots enable more natural interaction with users and can provide personalized recommendations through Natural Language Processing (NLP) technology. This study aims to develop a chatbot-based tourism recommendation system capable of delivering personalized destination suggestions based on user preferences. The system leverages a combination of approaches: Collaborative Filtering (CF) using the SVD++ algorithm, Content-Based Filtering (CBF) based on TF-IDF and cosine similarity, and a Knowledge-Based Recommender that filters results based on category, location, and minimum desired rating. To enhance flexibility and accuracy, a Hybrid Filtering approach is applied using a weighted hybrid mechanism that merges predictions from both CF and CBF with adjustable weights. The system also implements a switching strategy to select the most appropriate recommendation model based on the user's interaction history. Model evaluation results show that the hybrid approach achieved an RMSE of 0.6041 and an MAE of 0.4185, outperforming the individual CF and CBF models. Furthermore, the integration with a chatbot using the RASA Framework enables users to interact in real time through conversational interfaces to obtain relevant tourism recommendations. The results of this study indicate that the proposed system is not only capable of addressing cold-start problems but also provides a more responsive and contextual user experience.

Keywords: *Recommender System, Chatbot, Hybrid Filtering, Collaborative Filtering, Content-Based Filtering, RASA Framework*

DAFTAR ISI

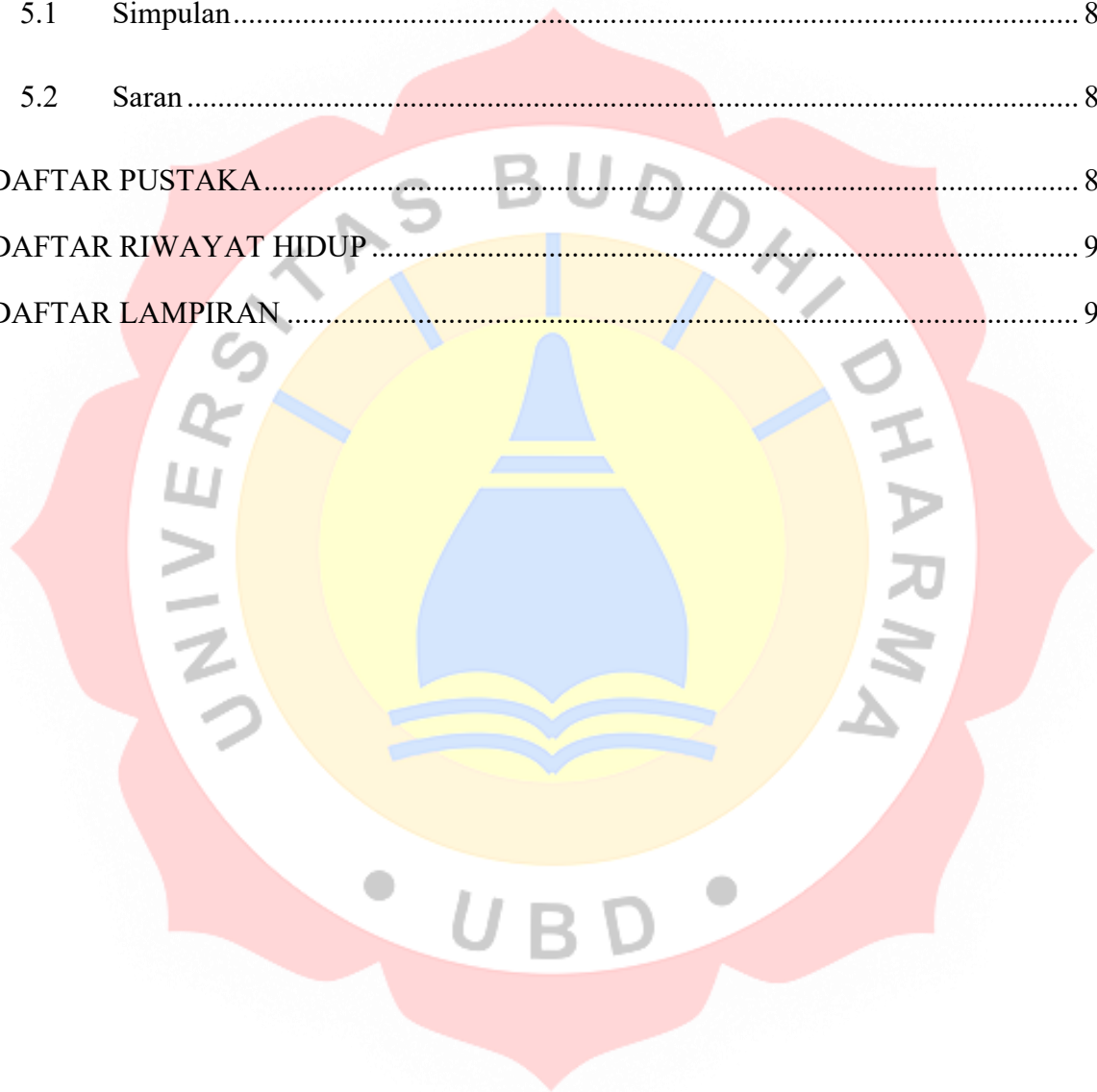
LEMBAR PERNYATAAN KEASLIAN SKRIPSI.....	i
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH	ii
LEMBAR PENGESAHAN PEMBIMBING.....	iii
LEMBAR PENGESAHAN SKRIPSI.....	iv
LEMBAR PENGESAHAN TIM PENGUJI	v
KATA PENGANTAR.....	vi
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI	ix
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	4
1.3 Ruang Lingkup	4
1.4 Tujuan dan Manfaat.....	5
1.4.1 Tujuan.....	5
1.4.2 Manfaat.....	6
1.5 Sistematika Penulisan	6
BAB II LANDASAN TEORI.....	8
2.1 Data.....	8
2.2 <i>Data Mining</i>	8

2.3	<i>Aplikasi Mobile</i>	9
2.4	Kecerdasan Buatan	9
2.5	<i>Machine Learning</i>	10
2.6	Sistem Rekomendasi.....	10
2.7	<i>Chatbot</i>	11
2.8	<i>Natural Language Processing</i>	11
2.9	<i>Collaborative Filtering</i>	11
2.10	<i>Content-Based Filtering</i>	13
2.11	<i>Knowledge-Based Recommender</i>	13
2.12	<i>Popularity-Based Recommender</i>	15
2.13	<i>Hybrid Methods</i>	16
2.14	RMSE	17
2.15	MAE	18
2.16	<i>Precision, Recall, F1-Score</i>	19
2.17	<i>RASA Framework</i>	20
2.18	<i>Python</i>	20
2.19	<i>Flowchart</i>	21
2.20	Tinjauan Studi.....	22
2.21	Kesimpulan Tinjauan Studi	28
BAB III METODE PENELITIAN		30
3.1	Kerangka Pemikiran	30

3.2	<i>Data Collection</i>	32
3.3	<i>Data Understanding</i>	34
3.4	<i>Data Preprocessing</i>	36
3.4.1	<i>Data Cleaning</i>	36
3.4.2	<i>Data Filtering</i>	40
3.5	<i>Data Splitting</i>	41
3.6	<i>Model Development</i>	42
3.6.1	<i>Flowchart Model</i>	42
3.6.2	<i>Collaborative Filtering</i>	43
3.6.3	<i>Content-Based Filtering</i>	44
3.6.4	<i>Knowledge Based</i>	46
3.6.5	<i>Popularity-Based</i>	47
3.6.6	<i>Hybrid Approach</i>	48
3.7	<i>Metode Evaluasi</i>	49
3.7.1	<i>Mean Absolute Error</i>	49
3.7.2	<i>Root Mean Squared Error</i>	49
3.8	<i>Precision, Recall, F1-Score</i>	50
3.9	<i>Lingkungan Pengembangan</i>	50
3.9.1	<i>Spesifikasi Hardware</i>	51
3.9.2	<i>Spesifikasi Software</i>	51
3.10	<i>Chatbot Development</i>	52

3.10.1	Arsitektur dan Teknologi.....	52
3.10.2	Deteksi <i>Intent</i> dan Entitas.....	53
3.10.3	Aksi Kustom dan Integrasi API.....	54
3.10.4	<i>Rules dan Story</i>	54
BAB IV HASIL DAN PEMBAHASAN.....		56
4.1	Evaluasi Model.....	56
4.1.1	Evaluasi <i>Collaborative Filtering</i>	56
4.1.2	Evaluasi <i>Content-Based Filtering</i>	58
4.1.3	Evaluasi <i>Knowledge-Based</i>	59
4.1.4	Evaluasi <i>Popularity-Based</i>	60
4.1.5	Evaluasi <i>Hybrid</i>	61
4.2	Implementasi <i>Chatbot</i>	62
4.2.1	Fungsi Modular CF.....	63
4.2.1	Fungsi Modular CBF.....	65
4.2.2	Fungsi Modular <i>Hybrid</i>	67
4.2.1	Fungsi Modular KBSR.....	69
4.2.2	Fungsi Modular PBSR.....	71
4.2.3	<i>RestAPI</i>	73
4.3	Tampilan Aplikasi.....	74
4.3.1	Halaman Utama.....	74
4.3.2	Halaman <i>Search</i>	77

4.3.3	Halaman <i>Chatbot</i>	79
4.3.4	Halaman <i>Account</i>	81
4.4	Pembahasan	82
BAB V SIMPULAN DAN SARAN		85
5.1	Simpulan.....	85
5.2	Saran	86
DAFTAR PUSTAKA.....		88
DAFTAR RIWAYAT HIDUP		93
DAFTAR LAMPIRAN.....		94



DAFTAR GAMBAR

Gambar 2.1 Prinsip <i>Matrix Factorization</i>	12
Gambar 3.1 Kerangka Pemikiran	30
Gambar 3.2 Halaman input pada alat <i>google maps scrapper/extractor</i>	33
Gambar 3.3 Halaman hasil pada alat <i>google maps scrapper/extractor</i>	33
Gambar 3.4 Flowchart Model	42
Gambar 3.5 <i>Visual Studio Code</i>	52
Gambar 4.1 Halaman Utama	75
Gambar 4.2 Halaman detail wisata	76
Gambar 4.3 Halaman <i>Search</i>	78
Gambar 4.4 Halaman <i>Chatbot</i>	80
Gambar 4.5 Halaman <i>Account</i>	81

DAFTAR TABEL

Tabel 2.1 Simbol <i>Flowchart</i>	21
Tabel 3.1 Data Ringkasan Tempat Wisata	34
Tabel 3.2 Data Ulasan Detail.....	35
Tabel 3.3 Informasi <i>Dataset</i> sebelum proses <i>cleaning</i>	36
Tabel 3.4 Informasi <i>Dataset</i> setelah proses <i>cleaning</i>	37
Tabel 3.5 Informasi <i>Dataset</i> wisata sebelum proses <i>cleaning</i>	38
Tabel 3.6 Informasi <i>Dataset</i> wisata sesudah proses <i>cleaning</i>	39
Tabel 3.7 <i>Mapping</i> kategorikal.....	40
Tabel 3.8 Tabel CF.....	44
Tabel 3.9 Tabel CBF	45
Tabel 3.10 Informasi <i>intent</i> dan entitas	47
Tabel 3.11 Spesifikasi Hardware.....	51
Tabel 4.1 Eksperimen parameter terbaik.....	56
Tabel 4.2 Evaluasi metrik SVD++	57
Tabel 4.3 Evaluasi CBF.....	58
Tabel 4.4 Evaluasi Knowledge-Based.....	59
Tabel 4.5 Evaluasi <i>Popularity-Based</i>	61
Tabel 4.6 Evaluasi Hybrid Filtering	62

DAFTAR LAMPIRAN

Kartu Bimbingan	94
-----------------------	----



BAB I

PENDAHULUAN

1.1 Latar Belakang

Sektor pariwisata memiliki peran penting dalam mendorong perekonomian suatu negara. Kontribusi pariwisata terhadap pertumbuhan ekonomi pada banyak negara dinilai cukup signifikan (UNWTO, 2023). Sebagai pendorong pertumbuhan ekonomi, pariwisata dapat menyediakan peluang kerja dan menjadi salah satu sumber pendapatan utama bagi negara (Kurniawan et al., 2024). Pariwisata berperan penting dalam memperkenalkan kebudayaan lokal, melestarikan warisan budaya, dan meningkatkan kualitas hidup masyarakat di sekitar destinasi wisata (Gameliel et al., 2024). Untuk meningkatkan dampak ekonomi dan mendukung peran pentingnya, sektor pariwisata dapat menggunakan bantuan teknologi (Aolia et al., 2024). Dalam hal ini, pemanfaatan teknologi dibutuhkan agar sektor pariwisata dapat mengikuti perkembangan di era digital.

Seiring dengan pesatnya perkembangan teknologi di era digital, sektor pariwisata dituntut untuk berinovasi dalam meningkatkan kepuasan wisatawan domestik dan internasional. Pemanfaatan teknologi seperti Kecerdasan Buatan atau yang dikenal dengan *Artificial Intelligence* (AI) dan platform digital kini menjadi bagian utama dari transformasi ini (Simoes et al., 2022). Penggunaan AI memungkinkan bagi wisatawan untuk mengakses informasi dan layanan yang lebih mudah (Zhang et al., 2022). AI telah membawa perubahan signifikan pada berbagai aspek kehidupan manusia, termasuk cara mereka bekerja, bergajian, dan berinteraksi. Dengan penerapan teknologi AI yang semakin luas, kebutuhan sektor pariwisata dalam proses transformasi digital akan membuka peluang besar untuk otomatisasi berbasis teknologi cerdas (Duarte et al., 2020).

Salah satu aplikasi teknologi kecerdasan buatan yang menjanjikan dalam sektor pariwisata adalah penggunaan *Chatbot AI*. *Chatbot* merupakan program komputer yang mampu meniru interaksi dengan manusia melalui teknologi *Natural Language Processing* (NLP) dan algoritma *Machine Learning* (ML) (Bird *et al.*, 2023). *Chatbot* ini dapat dikembangkan untuk merespon pertanyaan, memberikan saran, serta menyajikan pengalaman yang disesuaikan bagi setiap pengguna (Ali *et al.*, 2023). Dengan kemampuan *chatbot* untuk memahami dan merespon interaksi pengguna, teknologi ini juga dapat dipadukan dengan sistem rekomendasi yang memungkinkan penyajian informasi lebih personal dan relevan berdasarkan preferensi wisatawan (Ho *et al.*, 2023).

Sistem rekomendasi dapat membantu wisatawan dalam memilih destinasi, aktivitas, atau layanan berdasarkan preferensi pribadi, riwayat pencarian, dan data lainnya. Dalam konteks ini, algoritma sistem rekomendasi seperti *Collaborative Filtering*, *Content-Based Filtering*, dan *Hybrid Methods* telah digunakan untuk menganalisis data besar dan memberikan saran yang lebih tepat (Amri Akbar *et al.*, 2021). Dengan memanfaatkan data yang diperoleh dari interaksi pengguna, sistem rekomendasi dapat mempersonalisasi pengalaman perjalanan, memberikan pilihan destinasi yang sesuai, serta meningkatkan kepuasan wisatawan (H. T. Nguyen *et al.*, 2023).

Personalisasi dalam sektor pariwisata memiliki peran yang sangat penting, karena setiap wisatawan memiliki preferensi yang berbeda-beda, yang dipengaruhi oleh minat pribadi, lokasi, serta pengalaman perjalanan sebelumnya (Chanrueang *et al.*, 2024). Dengan memanfaatkan teknologi *chatbot* berbasis AI dan sistem rekomendasi, sistem dapat mempelajari pola perilaku pengguna melalui interaksi mereka dan memberikan rekomendasi yang lebih tepat serta relevan. Hal ini tidak hanya meningkatkan relevansi pengalaman wisatawan, tetapi juga berpotensi untuk meningkatkan tingkat kepuasan dan loyalitas pengguna terhadap layanan yang diberikan (Mei *et al.*, 2024). Oleh karena itu, integrasi

teknologi ini berkontribusi besar dalam menciptakan pengalaman yang lebih disesuaikan dan memuaskan bagi setiap individu dalam perjalanan mereka.

Integrasi *chatbot* AI dan sistem rekomendasi dapat menghadirkan solusi yang lebih cerdas untuk membantu wisatawan merencanakan perjalanan (Cassani *et al.*, 2024). Chatbot berperan sebagai antarmuka percakapan interaktif yang mengumpulkan informasi wisatawan, lalu memberikan rekomendasi *real-time* yang personal dan responsif (Rosida & Hadiono, 2024). Dilengkapi kemampuan rekomendasi, *chatbot* juga dapat memberikan respons kontekstual, seperti menyarankan destinasi sesuai minat pengguna atau menawarkan layanan tambahan berdasarkan kebutuhan yang teridentifikasi selama percakapan (Lin *et al.*, 2023). Dengan demikian, kombinasi teknologi *chatbot* dan sistem rekomendasi menjadi alat efektif untuk memperkaya pengalaman wisatawan sekaligus memudahkan pengambilan keputusan selama perjalanan.

Berdasarkan latar belakang yang telah diuraikan di atas, penelitian ini berfokus pada pengembangan sistem rekomendasi berbasis *chatbot* AI yang menggabungkan metode *Collaborative Filtering* dan *Content-Based Filtering* dalam pendekatan *hybrid*. Sistem ini dirancang untuk memberikan rekomendasi tempat wisata yang relevan berdasarkan preferensi pengguna, *rating* positif, dan lokasi, dengan memanfaatkan *dataset* primer yang mencakup wilayah Bali, Malang, Yogyakarta, Bandung, dan Jakarta dengan masing-masing 10 tempat wisata terkenal di kota tersebut. Selain itu, penelitian ini juga mencakup pengembangan aplikasi *chatbot* berbasis *mobile* untuk mempermudah wisatawan dalam mendapatkan rekomendasi yang sesuai dengan kebutuhan mereka. Dengan demikian, penelitian ini berjudul **"CHATBOT BERBASIS KECERDASAN BUATAN DENGAN SISTEM REKOMENDASI HYBRID UNTUK REKOMENDASI DESTINASI WISATA"**

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan, masalah yang dapat diidentifikasi adalah sebagai berikut:

- a. Dibutuhkan solusi berbasis teknologi untuk memberikan rekomendasi tempat wisata yang relevan dengan preferensi pengguna.
- b. Diperlukan sistem rekomendasi untuk pengguna baru yang tidak memiliki riwayat interaksi.
- c. Kurangnya sistem rekomendasi yang dapat menggabungkan berbagai faktor (tema, *rating*, kesukaan pengguna, dll.) untuk memberikan rekomendasi yang lebih akurat dan personal.

1.3 Ruang Lingkup

Ruang lingkup penelitian ini adalah sebagai berikut:

- a. *Dataset* yang digunakan merupakan data hasil *scrapping* data ulasan *Google Maps* terhadap berbagai tempat di Indonesia *dataset* ini terdiri dari dua bagian utama, yaitu *overview* data yang mencakup informasi umum seperti, nama wisata, deskripsi, jumlah *rating*, dan jumlah ulasan, serta *detailed reviews* yang memuat data *granular* seperti data user, *rating* yang diberikan, teks ulasan, dan lain sebagainya.
- b. *Dataset* terdiri dari beberapa tabel yang mencakup informasi mengenai pengguna, ulasan, atraksi wisata, lokasi, dan jenis atraksi. Namun, *dataset* ini memiliki limitasi geografis karena hanya mencakup lima kota populer di Indonesia, yaitu Bali, Malang, Yogyakarta, Bandung, dan Jakarta. Setiap kota hanya mencakup 10 wisata terpopuler berdasarkan data ulasan yang dikumpulkan.

- c. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi tempat wisata berbasis *chatbot* AI yang mampu memberikan saran personal berdasarkan masukan pengguna seperti lokasi, *rating*, ulasan positif, dan preferensi tertentu yang dapat ditangkap dari percakapan.
- d. Pengembangan sistem mencakup implementasi *chatbot* interaktif berbasis *mobile* yang dapat memahami dan merespons pertanyaan seputar rekomendasi wisata, serta memanfaatkan informasi dari data ulasan publik sebagai sumber utama.
- e. Sistem rekomendasi yang dikembangkan mampu memberikan saran wisata dari berbagai wilayah di Indonesia selama tempat tersebut tersedia di dalam *dataset*. Namun, cakupan rekomendasi tetap dibatasi oleh entitas yang telah terdata dalam hasil *scrapping* dan tidak dapat mencakup lokasi yang belum terdokumentasi.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

- a. Mengembangkan sistem rekomendasi tempat wisata berbasis *chatbot* AI yang mampu memberikan saran destinasi wisata berdasarkan tema atau preferensi yang dimasukkan pengguna, seperti kategori atau lokasi.
- b. Menerapkan algoritma rekomendasi agar sistem tetap dapat memberikan rekomendasi yang relevan, bahkan untuk pengguna baru yang belum memiliki riwayat interaksi atau penilaian sebelumnya.
- c. Mengintegrasikan modul sistem rekomendasi dengan *chatbot* sehingga pengguna dapat berinteraksi secara alami dan menerima rekomendasi wisata, berdasarkan input teks maupun konteks percakapan.

1.4.2 Manfaat

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

- a. Pengguna mendapatkan rekomendasi tempat wisata yang relevan dengan tema yang dimasukkan, sehingga membantu mereka menemukan destinasi sesuai preferensi.
- b. Sistem membantu pengguna baru yang belum memiliki riwayat perjalanan dengan tetap memberikan rekomendasi yang relevan.
- c. Pengalaman pengguna meningkat berkat interaksi *real-time* melalui *chatbot*, yang menyajikan informasi dengan cepat dan responsif.

1.5 Sistematika Penulisan

Penulisan sistematis dimaksudkan untuk mempermudah pemahaman dan penelaahan penelitian. Dalam laporan penelitian ini, bahan yang ditulis dikelompokkan menjadi beberapa sub bab, yang secara garis besar dapat dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Bab ini memuat penjelasan mengenai latar belakang masalah, identifikasi masalah, ruang lingkup penelitian, tujuan dan manfaat penelitian, serta sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini memuat paparan mengenai teori-teori yang berhubungan dengan penelitian ini, yang terbagi menjadi 4 bagian, di antaranya adalah teori umum, teori khusus, teori rancangan, dan teori pengujian, serta mencakup tinjauan studi.

BAB III METODE PENELITIAN

Bab ini memuat paparan terkait metode yang digunakan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memuat hasil evaluasi terhadap model yang diangkat, tampilan program, dan pengujian aplikasi.

BAB V SIMPULAN DAN SARAN

Bab ini memuat penjelasan mengenai kesimpulan penelitian berdasarkan hasil evaluasi dan penerapan yang telah dilakukan, serta rekomendasi yang dapat dikembangkan untuk penelitian di masa mendatang.



BAB II

LANDASAN TEORI

2.1 Data

Data merujuk pada sekumpulan fakta, informasi, atau hasil pengamatan yang disajikan dalam beragam bentuk, seperti interval, nominal, ordinal, maupun persentase, dan dapat digunakan untuk mendukung kegiatan penelitian, pengambilan keputusan, serta berbagai keperluan lainnya (Umar Hussain, 2023). Secara umum, data dikategorikan ke dalam tiga tipe utama: data tidak terstruktur, data terstruktur, dan data semi-terstruktur. Data tidak terstruktur adalah jenis data yang tidak mengikuti pola tertentu, sehingga tidak dapat diakses melalui *query*. Sementara itu, data terstruktur dapat diakses oleh manusia maupun mesin dengan menggunakan *query*. Di antara keduanya, terdapat data semi-terstruktur yang memiliki pola fleksibel, seperti pada format JSON dan XML (Breitinger & Jotterand, 2023).

2.2 Data Mining

Data mining merupakan proses penyaringan atau penggalian informasi dari himpunan data dalam jumlah besar melalui sejumlah tahapan untuk memperoleh wawasan yang bernilai (Asyuti & Setyawan, 2023). Proses ini mencakup langkah-langkah seperti pengumpulan data, *pre-processing*, pembagian data menjadi data latih dan data uji, serta analisis dengan pendekatan *machine learning*. Hasil dari proses ini dapat berupa pola, keterkaitan, atau kecenderungan baru yang berguna dalam mendukung proses pengambilan keputusan. Lebih jauh, *data mining* adalah teknologi yang mampu memprediksi tren di masa mendatang, sehingga dapat membantu pengguna dalam membuat keputusan yang lebih tepat (Salem *et al.*, 2022).

2.3 Aplikasi Mobile

Aplikasi mobile adalah perangkat lunak yang dirancang untuk beroperasi pada perangkat *mobile* seperti *smartphone* dan *tablet*. Aplikasi ini umumnya terbagi menjadi dua kategori utama: *native apps*, yaitu aplikasi yang dikembangkan secara khusus untuk suatu platform tertentu (contohnya, *Android* atau *iOS*) dengan menggunakan bahasa pemrograman dan alat pengembangan yang sesuai; serta *cross-platform apps*, yang memungkinkan pembuatan aplikasi untuk beberapa platform hanya dengan satu basis kode (Mascetti *et al.*, 2021). Platform seperti *Flutter*, *React Native*, dan *Android Studio* mendukung pengembangan lintas platform secara lebih efisien dan menghemat waktu, karena hanya memerlukan satu bahasa pemrograman untuk menghasilkan aplikasi yang kompatibel dengan dua platform utama, yakni *Android* dan *iOS*. Saat ini, *aplikasi mobile* telah menjadi bagian integral dari kehidupan sehari-hari, memberikan kemudahan dalam mengakses informasi, layanan, dan berbagai fungsi lainnya termasuk dalam sektor pariwisata, di mana *aplikasi mobile* memungkinkan wisatawan memperoleh rekomendasi, informasi, dan layanan secara *real-time* (Hu & You, 2021).

2.4 Kecerdasan Buatan

Kecerdasan Buatan (AI) mengacu pada kapabilitas sistem komputer dalam menjalankan tugas-tugas yang umumnya membutuhkan kecerdasan manusia, seperti pemrosesan bahasa alami, pengenalan gambar, pengambilan keputusan, serta kemampuan belajar. AI mencakup beragam pendekatan, termasuk *machine learning*, yaitu metode di mana sistem dilatih untuk mengidentifikasi pola dari data, serta *deep learning*, yang memanfaatkan jaringan saraf tiruan untuk menangani informasi yang lebih kompleks. Teknologi ini semakin luas penerapannya di berbagai sektor, seperti kesehatan, pendidikan, dan pariwisata, guna meningkatkan efisiensi serta menghadirkan solusi yang inovatif (Muchlinski, 2022).

2.5 *Machine Learning*

Machine learning adalah salah satu bidang dalam *kecerdasan buatan* yang memungkinkan komputer untuk belajar dari pengalaman, serta mengidentifikasi pola dan keterkaitan antara *input* dan *output* dalam data melalui pendekatan *statistik* dan *probabilistik* yang beragam (Chowdhury *et al.*, 2021). *Machine learning* umumnya diklasifikasikan ke dalam dua kategori utama, yaitu *supervised learning* dan *unsupervised learning* (Dogan & Birant, 2021). Perbedaan mendasar antara keduanya terletak pada pendekatan pembelajarannya: *supervised learning* mempelajari relasi antara *input* dan *output* dengan menggunakan data berlabel, sedangkan *unsupervised learning* menganalisis data tanpa label secara mandiri, tanpa campur tangan manusia.

2.6 *Sistem Rekomendasi*

Sistem rekomendasi merupakan teknologi yang dirancang untuk menyajikan saran atau rekomendasi kepada pengguna berdasarkan data yang diperoleh dari riwayat interaksi, preferensi, atau perilaku pengguna sebelumnya (Didi Riswan *et al.*, 2024). Terdapat tiga pendekatan utama dalam penerapan sistem rekomendasi, yaitu *Collaborative Filtering*, *Content-Based Filtering*, dan *Hybrid Methods*. *Collaborative Filtering* memanfaatkan pola perilaku pengguna untuk menghasilkan rekomendasi dengan mengidentifikasi kemiripan dengan pengguna lain. Di sisi lain, *Content-Based Filtering* berfokus pada karakteristik *item* dan menyusun rekomendasi yang relevan berdasarkan ketertarikan pengguna terhadap fitur-fitur tertentu dari *item* tersebut (Mehrabani *et al.*, 2020). Sementara itu, *Hybrid Methods* menggabungkan kedua pendekatan tersebut untuk mengatasi keterbatasan masing-masing, seperti masalah *cold-start* pada *Collaborative Filtering*, sekaligus meningkatkan ketepatan rekomendasi (Trabelsi *et al.*, 2021).

2.7 Chatbot

Chatbot merupakan sistem yang berbasis pada *kecerdasan buatan* (AI) dan dirancang untuk meniru interaksi percakapan manusia melalui antarmuka teks maupun suara. Teknologi ini memanfaatkan *Natural Language Processing* (NLP) dan *Machine Learning* (ML) guna memahami, mengolah, serta memberikan respons terhadap masukan pengguna secara otomatis (Bird *et al.*, 2023). Secara umum, *chatbot* terbagi menjadi dua tipe utama: *rule-based chatbots*, yang beroperasi dengan mengikuti seperangkat aturan yang telah ditentukan sebelumnya, dan *AI-based chatbots*, yang memanfaatkan model *pembelajaran mesin* untuk terus meningkatkan kemampuan percakapan seiring waktu (Khenouche *et al.*, 2024).

2.8 Natural Language Processing

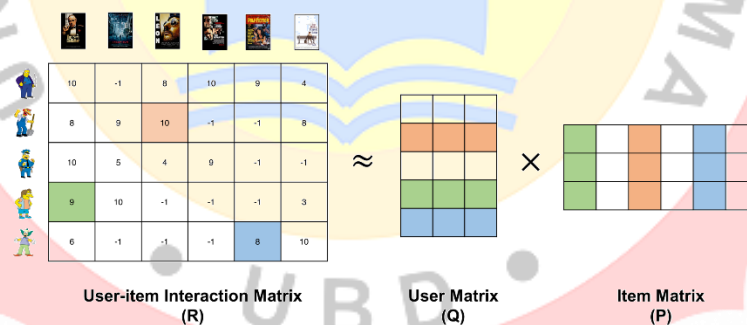
Natural Language Processing (NLP) merupakan salah satu cabang utama dalam *kecerdasan buatan* yang berfokus pada hubungan antara komputer dan bahasa manusia. NLP memungkinkan sistem untuk memahami, menganalisis, serta menghasilkan teks atau ucapan manusia secara otomatis. Teknologi ini memainkan peran penting dalam berbagai aplikasi, seperti *chatbot*, penerjemah otomatis, dan analisis sentimen semuanya bergantung pada kemampuan mesin dalam memproses bahasa alami. NLP mencakup beragam teknik, seperti *tokenization*, *stemming*, *named entity recognition* (NER), dan *part-of-speech tagging*, yang digunakan untuk mengekstraksi informasi dari teks agar dapat dimengerti oleh mesin (Lan *et al.*, 2024).

2.9 Collaborative Filtering

Collaborative Filtering (CF) merupakan salah satu pendekatan utama dalam sistem rekomendasi yang digunakan untuk memprediksi preferensi pengguna berdasarkan perilaku atau kesamaan preferensi dengan pengguna lain. CF terbagi ke dalam dua pendekatan utama: *User-Based Collaborative Filtering* dan *Item-Based Collaborative Filtering*. Pada *User-*

Based CF, sistem menyusun rekomendasi dengan cara mencari pengguna lain yang memiliki preferensi serupa, lalu menyarankan *item* yang telah disukai oleh pengguna-pengguna tersebut (Huang *et al.*, 2020). Sebaliknya, *Item-Based CF* berfokus pada analisis kesamaan antar *item* berdasarkan interaksi pengguna, kemudian merekomendasikan *item* yang mirip dengan yang sebelumnya dinilai positif oleh pengguna (Yusmar *et al.*, 2021).

Matrix Factorization adalah salah satu teknik yang banyak diterapkan dalam pengembangan sistem rekomendasi berbasis *Collaborative Filtering*. Pada Gambar 2.1 dijelaskan bahwa teknik ini bertujuan untuk memperkirakan preferensi pengguna terhadap suatu *item* dengan mengidentifikasi pola interaksi yang tercermin dalam *user-item matrix* (Huang *et al.*, 2020). Salah satu metode *Matrix Factorization* yang cukup populer adalah *Singular Value Decomposition* (SVD++), yang bekerja dengan memecah *matriks* besar menjadi beberapa komponen yang lebih kecil dan dapat diolah untuk menghasilkan rekomendasi yang lebih personal dan akurat (Peng *et al.*, 2022).



Gambar 2.1 Prinsip *Matrix Factorization*

(Khan *et al.*, 2023)

Pada sistem rekomendasi, *matriks user-item* biasanya memiliki dimensi besar dan sparsitas tinggi, yang berarti sebagian besar elemen dalam *matriks* tersebut bernilai nol akibat minimnya interaksi pengguna terhadap sebagian besar *item*. Dengan SVD++, *matriks user-item R* didekomposisi menjadi tiga *matriks* kecil:

$$R \approx U\Sigma V^T \quad (1)$$

di mana:

- a. U adalah *matriks* yang merepresentasikan pengguna dalam ruang fitur *laten*.
- b. Σ adalah *matriks* diagonal yang merepresentasikan bobot atau *singular value* dari fitur *laten*.
- c. V^T adalah *matriks* yang merepresentasikan *item* dalam ruang fitur *laten*.

2.10 Content-Based Filtering

Content-Based Filtering (CBF) merupakan salah satu metode dalam sistem rekomendasi yang menyajikan saran kepada pengguna berdasarkan karakteristik atau atribut dari *item* yang telah dilihat atau disukai oleh pengguna tersebut. Dalam pendekatan ini, setiap *item* memiliki sejumlah deskripsi atau fitur yang diekstraksi dan dianalisis untuk mengidentifikasi kemiripan dengan *item* lainnya. Sistem kemudian akan merekomendasikan *item* yang memiliki atribut serupa dengan *item* yang sebelumnya diberi nilai positif oleh pengguna (Mehrabani *et al.*, 2020). Sebagai contoh, dalam konteks pariwisata, sistem CBF dapat memanfaatkan deskripsi atau kategori destinasi wisata (seperti pantai, pegunungan, atau budaya) untuk merekomendasikan tempat wisata lain yang memiliki tema serupa. Kekuatan utama dari CBF terletak pada kemampuannya memberikan rekomendasi yang sangat relevan dan personal, disesuaikan dengan preferensi unik masing-masing pengguna. Meski demikian, pendekatan ini juga memiliki keterbatasan, seperti kurangnya variasi dalam hasil rekomendasi serta ketergantungan terhadap kelengkapan dan kualitas data fitur yang tersedia (Javed *et al.*, 2021).

2.11 Knowledge-Based Recommender

Knowledge-Based Recommender System (KBRS) adalah sistem rekomendasi yang menghasilkan saran berdasarkan aturan eksplisit, profil pengguna, dan atribut *item* yang

dapat diobservasi secara langsung, tanpa memerlukan riwayat interaksi pengguna sebelumnya. Sistem ini beroperasi dengan mencocokkan preferensi atau kebutuhan pengguna seperti kategori *item*, lokasi, dan batas minimum *rating* dengan atribut yang tersedia pada *item* di basis data, menggunakan pendekatan *rule-based* atau *constraint-based reasoning*. Tidak seperti CF yang bergantung pada kesamaan pengguna atau *item* dari interaksi historis, KBRS cocok untuk skenario *cold-start*, di mana pengguna baru atau *item* baru belum memiliki cukup riwayat interaksi. Dalam sistem ini, pengambilan keputusan didasarkan pada *filtering* eksplisit, yaitu menyaring *item* yang tidak sesuai dengan preferensi pengguna dan hanya mengembalikan yang relevan.

Menurut penelitian terbaru oleh (Flórez *et al.*, 2024), KBRS menjadi sangat relevan dalam sistem rekomendasi wisata karena pengguna cenderung memiliki kebutuhan eksplisit, seperti preferensi destinasi berdasarkan lokasi geografis, tipe wisata, atau tingkat kenyamanan. Mereka menunjukkan bahwa integrasi KBRS ke dalam sistem rekomendasi dapat membantu meningkatkan relevansi rekomendasi awal, serta memperkaya skenario interaksi pada sistem hybrid (Flórez *et al.*, 2024).

Selain itu, penelitian oleh (Uta *et al.*, 2024) menunjukkan bahwa KBRS juga unggul dalam hal transparansi dan pengendalian preferensi, di mana pengguna dapat secara langsung melihat alasan di balik setiap rekomendasi yang diberikan—karena rekomendasi dibentuk berdasarkan aturan yang eksplisit, bukan korelasi statistik tersembunyi. Namun, KBRS juga memiliki tantangan tersendiri, seperti kebutuhan akan desain aturan yang kompleks serta keterbatasan eksplorasi *item* alternatif jika aturan terlalu ketat. Oleh karena itu, pendekatan ini sering dipadukan dalam kerangka *hybrid recommender*, agar tetap dapat menangani eksplorasi berbasis konten maupun data historis.

2.12 *Popularity-Based Recommender*

Sistem rekomendasi berbasis popularitas (*Popularity-Based Recommender System*) atau PBRS merupakan salah satu pendekatan paling sederhana dalam *domain* sistem rekomendasi. Prinsip kerja metode ini adalah memberikan rekomendasi kepada pengguna berdasarkan tingkat popularitas suatu *item*, yang diukur dari metrik tertentu seperti jumlah pembelian, jumlah unduhan, jumlah penayangan, atau rata-rata rating tertinggi pada periode waktu tertentu (Jing *et al.*, 2023). Popularitas ini dihitung secara agregat dan tidak mempertimbangkan profil atau preferensi individu pengguna. Dengan kata lain, seluruh pengguna akan mendapatkan daftar rekomendasi yang sama, yaitu *item* yang sedang populer di kalangan pengguna lain.

Keunggulan utama dari metode ini terletak pada kesederhanaannya. *Popularity-based recommender* tidak memerlukan data personal yang kompleks ataupun pemrosesan berbasis *machine learning* yang berat. Algoritma ini hanya membutuhkan proses peringkatan berdasarkan frekuensi atau tingkat interaksi terhadap *item*, sehingga mampu memberikan hasil rekomendasi dengan cepat dan biaya komputasi yang rendah (Ji *et al.*, 2020). Hal ini membuatnya sering digunakan sebagai *baseline* atau pembanding dalam penelitian sistem rekomendasi, serta sebagai metode *fallback* pada skenario *cold-start*.

Namun, metode ini memiliki keterbatasan signifikan, terutama terkait *popularity bias*. Bias ini muncul ketika sistem secara berulang merekomendasikan *item* yang sudah populer, sehingga mengurangi peluang eksposur *item* yang kurang dikenal. Fenomena ini sering disebut efek “*rich-get-richer*”, di mana *item* populer semakin mendominasi dan *item* baru atau *niche* kesulitan mendapatkan visibilitas (Zhang *et al.*, 2021). Selain itu, rekomendasi berbasis popularitas bersifat *non-personalized*, sehingga kualitas pengalaman pengguna yang menginginkan rekomendasi sesuai preferensi pribadinya dapat menurun.

Secara umum, meskipun metode *popularity-based* tidak dapat sepenuhnya menggantikan algoritma yang lebih kompleks dalam sistem rekomendasi modern, keberadaannya tetap relevan. *Popularity-based recommender* berperan sebagai *baseline*, metode *fallback* pada *cold-start*, dan sebagai komponen dalam model *hybrid* yang memanfaatkan keunggulannya dalam kesederhanaan, kecepatan, serta kemampuannya menangkap tren global secara efisien.

2.13 Hybrid Methods

Hybrid Methods dalam sistem rekomendasi adalah pendekatan yang menggabungkan dua atau lebih metode rekomendasi, seperti *Collaborative Filtering* (CF) dan *Content-Based Filtering* (CBF), untuk mengatasi kekurangan masing-masing metode dan meningkatkan akurasi rekomendasi (Khennouche *et al.*, 2024). Pendekatan ini bertujuan untuk memanfaatkan keunggulan dari berbagai teknik untuk menghasilkan sistem rekomendasi yang lebih kuat dan fleksibel. Misalnya, *Hybrid Methods* dapat mengatasi masalah *cold-start* yang umum dalam *Collaborative Filtering* dengan menggabungkannya dengan *Content-Based Filtering* yang menggunakan atribut *item* untuk memberikan rekomendasi, bahkan tanpa riwayat interaksi pengguna (Prathima *et al.*, 2020). Berbagai teknik yang digunakan dalam *Hybrid Methods* termasuk *weighted hybrid*, *switching hybrid*, dan *mixed hybrid*, di mana masing-masing memiliki cara berbeda dalam menggabungkan informasi dari berbagai sumber untuk menghasilkan rekomendasi yang lebih relevan dan personal (Trabelsi *et al.*, 2021).

Pada sistem *Weighted Hybrid*, setiap teknik rekomendasi menghasilkan skor yang memprediksi preferensi pengguna terhadap suatu *item*. Skor ini kemudian digabungkan menjadi satu skor komposit menggunakan bobot tertentu.

$$S(u, i) = w_{CF} \times S_{CF}(u, i) + w_{CBF} \times S_{CBF}(u, i) \quad (2)$$

Di mana:

- $S(u, i)$: adalah skor gabungan untuk pengguna u dan *item* i .
- $S_{CF}(u, i)$: Skor yang dihasilkan oleh *Collaborative Filtering*.
- $S_{CBF}(u, i)$: Skor yang dihasilkan oleh *Content-Based Filtering*.
- w_{CF} : Bobot untuk *Collaborative Filtering*.
- w_{CBF} : Bobot untuk *Content-Based Filtering*.

Bobot w_{CF} dan w_{CBF} dipilih sedemikian rupa sehingga $w_{CF} + w_{CBF} = 1$. Dengan menyesuaikan bobot, sistem dapat mengatur seberapa besar pengaruh setiap metode dalam memberikan rekomendasi.

2.14 RMSE

Root Mean Square Error (RMSE) merupakan salah satu metrik evaluasi paling umum digunakan untuk mengukur seberapa besar perbedaan antara nilai yang diprediksi oleh sistem dan nilai aktual (*ground truth*). RMSE sangat relevan dalam sistem rekomendasi berbasis prediksi *rating*, karena mampu memberikan ukuran kesalahan dalam satuan yang sama dengan skala *rating* itu sendiri. Secara formal, RMSE didefinisikan sebagai akar kuadrat dari rata-rata kuadrat selisih antara nilai prediksi ($\hat{r}_{ui}$) dan nilai aktual (r_{ui}) untuk setiap pasangan pengguna *item* (u, i) . Rumusnya dituliskan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{N} \sum_{(u,i) \in T} (r_{ui} - \hat{r}_{ui})^2} \quad (3)$$

Di mana:

- N adalah jumlah total pasangan *user-item* dalam data uji T ,
- r_{ui} adalah *rating* aktual dari pengguna u terhadap *item* i ,
- $\hat{r}_{ui}$ adalah *rating* yang diprediksi oleh sistem.

Nilai RMSE yang lebih kecil menunjukkan prediksi model lebih akurat. RMSE memberikan penalti lebih besar terhadap kesalahan yang ekstrem karena menggunakan kuadrat dari error,

sehingga lebih sensitif terhadap outlier dibanding metrik lainnya seperti MAE (*Mean Absolute Error*) (T. T. Nguyen *et al.*, 2021).

2.15 MAE

Mean Absolute Error (MAE) merupakan salah satu metrik evaluasi yang digunakan untuk mengukur rata-rata besar kesalahan absolut antara nilai prediksi dan nilai aktual. Dalam konteks sistem rekomendasi, MAE digunakan untuk mengetahui seberapa jauh prediksi sistem terhadap *rating* pengguna menyimpang dari nilai sebenarnya, tanpa mempertimbangkan arah dari kesalahan tersebut. Secara matematis, MAE dirumuskan sebagai:

$$MAE = \frac{1}{N} \sum_{(u,i) \in T} |r_{ui} - \hat{r}_{ui}| \quad (4)$$

Di mana:

- N adalah jumlah pasangan *user-item* dalam data uji T .
- r_{ui} adalah nilai *rating* aktual yang diberikan pengguna u pada *item* i .
- $\hat{r}_{ui}$ adalah nilai *rating* yang diprediksi oleh sistem.

MAE memberikan gambaran seberapa besar kesalahan prediksi secara rata-rata dalam skala *rating*, dan lebih toleran terhadap outlier dibandingkan RMSE karena tidak menggunakan kuadrat dari error (A. Alsalemi *et al.*, 2020). MAE dianggap sebagai metrik yang intuitif dan mudah diinterpretasikan, sehingga banyak digunakan dalam evaluasi awal sistem rekomendasi, terutama ketika distribusi *rating* bersifat seragam atau tidak mengandung banyak nilai ekstrem.

2.16 Precision, Recall, F1-Score

Precision@K merupakan metrik evaluasi yang digunakan untuk mengukur sejauh mana hasil rekomendasi *top-K* mengandung *item* yang relevan bagi pengguna. *Precision@K* dihitung sebagai proporsi *item* relevan dari seluruh *item* yang direkomendasikan sebanyak *K*. Rumus *Precision@K* dituliskan sebagai:

$$Precision@K = \frac{|rekomendasi\ relevan \cap rekomendasi\ top-K|}{K} \quad (5)$$

Nilai *Precision@K* tinggi menunjukkan bahwa sebagian besar *item* yang direkomendasikan benar-benar sesuai dengan preferensi pengguna. Metrik ini sangat berguna dalam sistem di mana hanya sejumlah kecil *item* ditampilkan kepada pengguna (Yusmar *et al.*, 2021).

Recall@K digunakan untuk mengukur seberapa besar proporsi *item* relevan yang berhasil ditangkap oleh sistem dalam daftar *top-K* rekomendasi. Dengan kata lain, *recall* menunjukkan cakupan dari sistem rekomendasi terhadap *item* relevan yang tersedia. Rumus *Recall@K*:

$$Recall@K = \frac{|rekomendasi\ relevan \cap rekomendasi\ top-K|}{|total\ item\ relevan|} \quad (6)$$

Nilai *Recall@K* yang tinggi menunjukkan bahwa sistem mampu mencakup sebagian besar *item* relevan bagi pengguna, meskipun tidak semua muncul di urutan teratas (Argina, 2020).

F1-Score@K merupakan harmonic mean dari *Precision@K* dan *Recall@K*, yang digunakan sebagai ukuran gabungan untuk menyeimbangkan keduanya. *F1-Score@K* sangat berguna ketika penting untuk mempertimbangkan presisi dan cakupan secara bersamaan dalam evaluasi performa sistem rekomendasi. Rumus *F1-Score@K*:

$$F1@K = 2 \times \frac{Precision@K \times Recall@K}{Precision@K + Recall@K} \quad (7)$$

F1@K idealnya digunakan ketika *Precision* dan *Recall* memiliki tingkat kepentingan yang setara dalam konteks aplikasi rekomendasi (T. T. Nguyen *et al.*, 2021).

2.17 RASA Framework

RASA Framework adalah salah satu platform open-source yang digunakan untuk membangun *chatbot* berbasis kecerdasan buatan dengan kemampuan pemahaman bahasa alami (*Natural Language Understanding/NLU*) dan pengelolaan *dialog* (*dialog management*). RASA memungkinkan pengembangan *chatbot* yang dapat berinteraksi dengan pengguna secara kontekstual melalui pemrosesan bahasa alami dan pembelajaran mesin (T. T. Nguyen *et al.*, 2021). Framework ini terdiri dari dua komponen utama: RASA *NLU*, yang bertanggung jawab untuk memahami maksud (*intent*) dan *entitas* dalam input pengguna, dan RASA *Core* yang mengelola alur percakapan berdasarkan konteks dan riwayat percakapan sebelumnya (Anindyati, 2023).

2.18 Python

Python adalah bahasa pemrograman yang bersifat terinterpretasi, interaktif, dan berorientasi objek, dengan kemampuan yang kuat berkat penggabungan modul, pengecualian, tipe data tingkat tinggi, dan kelas. *Python* memiliki sintaks yang sangat jelas, sehingga mudah dipahami (Python Software Foundation, 2024). Keunggulannya mencakup dukungan lintas platform, termasuk *Linux* yang umumnya menyertakan *Python* sebagai bagian standar. *Python* juga menyediakan beragam modul siap pakai, seperti *Matplotlib*, *Numpy*, *Pandas*, *Scikit-learn*, dan banyak pustaka lainnya yang mendukung berbagai kebutuhan pengembangan (Kurniawan *et al.*, 2024).

Python memiliki beragam kegunaan, termasuk pengembangan aplikasi web, pengelolaan alur kerja, integrasi dengan basis data, pembacaan dan modifikasi file, pemrosesan data dalam skala besar, pelaksanaan operasi matematika, serta mempercepat

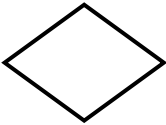
proses pembuatan prototipe dan pengembangan lebih lanjut perangkat lunak yang telah ada (Zein Hanni Pradana et al., 2022).

2.19 Flowchart

Tools berupa *flowchart* digunakan agar algoritma lebih mudah dipahami oleh orang lain, terutama programmer yang bertanggung jawab mengimplementasikan program. *Flowchart* adalah representasi grafis yang menggunakan simbol-simbol standar untuk menunjukkan serangkaian langkah atau proses dalam suatu sistem atau program (Chaudhuri, 2020). Setiap simbol dalam diagram alur mewakili proses tertentu, dan hubungan antar proses ditunjukkan dengan garis penghubung. Diagram alur membantu memperjelas dan memahami alur proses kegiatan. Tabel 2.1 mencantumkan simbol-simbol standar yang umum digunakan dalam diagram alur.

Tabel 2.1 Simbol *Flowchart*

Simbol	Name	Deskripsi
	Start/End	Menunjukkan Dimana sebuah proses dimulai atau berakhir
	Input/Output	Menandakan proses input dan output data
	<i>Process</i>	Menunjukkan proses yang dilakukan oleh komputer
	<i>Predefined Process</i>	Menunjukkan pelaksanaan suatu bagian (sub-program)
	<i>Flow Line</i>	Digunakan untuk menghubungkan simbol

	<i>Decision</i>	Digunakan untuk memberi nilai awal pada suatu variabel.
	<i>Preparation</i>	Digunakan untuk memberi nilai awal pada suatu variabel.
	<i>Database</i>	Digunakan untuk menyatakan database atau basis data.
	On-page Reference	Digunakan untuk menghubungkan flowchart pada halaman yang sama.
	On-page Reference	Digunakan untuk menghubungkan flowchart pada halaman yang berbeda.

(Chaudhuri, 2020)

2.20 Tinjauan Studi

Literatur mengenai pemanfaatan *artificial intelligence* (AI) untuk pariwisata dalam lima tahun terakhir menunjukkan lintasan riset yang kaya mulai dari ekstraksi pengetahuan berbasis *named-entity recognition* (NER), pengembangan *chatbot* cerdas, sampai rekayasa model rekomendasi hibrida yang memadukan *collaborative filtering* (CF), *content-based filtering* (CBF), dan *sentiment analysis*. Gelombang pertama ditandai oleh upaya memahami bahasa *domain* wisata secara lebih presisi. Di sinilah penelitian Dhomas Hatta Fudholi dan kolega memanfaatkan BERT untuk NER, memetakan 183507 token ke dalam label *heritage*, *natural*, dan *purpose* serta membukukan skor F1 0,80 melampaui BiLSTM-CRF klasik (Fudholi *et al.*, 2022). Keberhasilan tersebut menggarisbawahi pentingnya model bahasa kontekstual dalam

mengeksrak entitas destinasi secara akurat, sekaligus mengamplifikasi pengalaman “*clueless travelers*” yang tidak memiliki referensi awal.

Berbekal fondasi pemahaman bahasa tadi, rentetan studi berikutnya menaruh fokus pada *chatbot* sebagai antarmuka utama. Zein Hanni Pradana dan tim, misalnya, menerapkan kerangka RASA pada objek wisata Candi Prambanan. Dengan 21 kategori FAQ dan pipeline NLU berbasis *tokenisasi-featurisasi*, mereka mencapai akurasi 91 % dan *F1-score* 0,95, seraya menunjukkan bahwa otomasi tanya-jawab mampu menekan biaya operasional di masa pandemi (Pradana *et al.*, 2022). Pendekatan serupa tetapi berorientasi rute perjalanan digarap oleh Puspa Miladin Nuraida Safitri dkk. melalui *Random Forest Classifier*. Empirisnya, konfigurasi 200 pohon dan kedalaman 20 memberikan *F1-score* 96,16 % indikasi bahwa model korelasi fitur berlapis efektif menyeleksi sinyal relevan (Safitri *et al.*, 2023).

Transisi menuju *deep learning* yang lebih dalam terlihat pada kerangka berbasis GRU yang dikemukakan Giancarlo Sperli. Dengan arsitektur *micro-services* dan integrasi *knowledge graph* budaya, model Seq2Seq-GRU bukan hanya menorehkan akurasi 99,94 % melampaui LSTM, tetapi juga menurunkan beban tugas pengguna menurut NASA-TLX memperlihatkan sinergi antara efisiensi komputasi dan beban kognitif (Sperli, 2021). Jika penelitian Sperli menekankan warisan budaya Eropa, studi Reem Alotaibi dkk. memusatkan perhatian pada kota Jeddah. “*Smart Guidance*” nama *chatbot* tersebut mengompilasi 160 *intent* berbahasa Arab dan memprosesnya dalam 3–14 detik, dengan 90,5 % partisipan menyetujui kegunaan *chatbot* bagi wisatawan (Alotaibi *et al.*, 2020). Temuan ini memperkuat tesis bahwa bahasa lokal dan latensi rendah adalah kunci adopsi di lapangan.

Keunggulan *sequence-to-sequence* juga dikaji oleh Gurucharan Singh Saluja dan rekan yang *men-deploy* LSTM. Dengan 130 422 pasangan *query-response* dan

dukungan *Google Maps* serta *Wikipedia*, sistem merespons kueri rute aman atau penginapan murah dalam 0,35–0,55 detik jauh di bawah waktu pencarian manual via *search engine* (Saluja *et al.*, 2021). Dalam konteks Asia Tenggara, Hai Thanh Nguyen dkk. memadukan BERT dan RoBERTa dalam aplikasi *Android* berbasis *Flutter*; BERT menorehkan 90 % akurasi klasifikasi 7 319 sampel dan memperkaya fungsi *admin* mulai pelatihan model hingga pembaruan konten (Nguyen *et al.*, 2023). Sementara itu, Patrick Steven Siaw Chock Chiong dkk. menunjukkan bahwa *Telegram* dapat menjadi *front-end* ringan bagi wisatawan Mauritius; pengukuran *intent confusion matrix* bahkan memperlihatkan nihil kesalahan klasifikasi, menandakan konsistensi dialog (Chiong *et al.*, 2022).

Di sisi pemodelan bahasa *domain*, Veronika Arefieva dan Roman Egger memperkenalkan TourBERT model BERT yang dilatih *from scratch* pada 3,6 juta ulasan wisata. TourBERT melampaui BERT-Base baik dalam klasifikasi sentimen maupun pemetaan topik, sekaligus memperlihatkan kluster visual yang lebih terpisah pada *TensorBoard*, suatu bukti bahwa *pre-training domain-spesifik* meningkatkan representasi semantik (Arefieva & Egger, 2021).

Sejalan dengan kemajuan *chatbot*, skema rekomendasi destinasi turut berevolusi. Uribe Iraola dkk. mengombinasikan *multilayer perceptron* dan CF berbasis perkalian *matriks* dalam aplikasi seluler Peru. Hasil survei lapangan 92,5 % kemudahan pemakaian dan 91,7 % peningkatan pengalaman mengesankan bahwa AI mampu mendorong loyalitas pengguna hingga 96,7 % (Iraola *et al.*, 2023). Choirul Huda dan tim kemudian mendorong batasan ini dengan model hibrida: UBCF, CBCF, *Demographic Filtering*, dan *aspect-based sentiment analysis* (ABSA). Regenerasi *matriks* sintesis menaikkan densitas *UI-Matrix* dan memangkas MAE sampai 84,7 % dibanding UBCF murni, menegaskan keampuhan taktik *matrix densification* dalam

meredam *cold-start* (Huda *et al.*, 2024). Di Indonesia, Danny Matthew Saputra dkk. mengeksplor ALS varian *matrix factorization* pada 10 000 interaksi *rating*. MAE 1,272 dan RMSE 1,475 diiterasi kelima menjadi titik optimum sebelum *overfitting*; regulasi $\lambda = 0,6$ teridentifikasi sebagai kunci stabilitas (Saputra *et al.*, 2024). Riset ini menyoroti pentingnya keseimbangan regulasi-iterasi dalam skenario data masif.

Penelitian Mario Casillo dkk. memperkaya dimensi konteks melalui *Context Dimension Tree* (CDT) yang memetakan hierarki informasi wisata dan memfasilitasi adaptasi narasi *real-time* via *LDA*. Akurasi 87,31 % dalam mendeteksi elemen konteks menandakan bahwa personalisasi narasi berbasis konteks efektif menghadirkan *story-telling* yang sesuai situasi lapangan (Casillo *et al.*, 2020). Lebih jauh, Yoan Amri Akbar dkk. membuktikan bahwa *weighted parallel hybrid* menggabungkan CB dan CF dengan SVD mencapai RMSE 0,5006 untuk destinasi Lombok, lebih rendah daripada metode tunggal, sekaligus membuka jalan mitigasi *sparsity* (Akbar *et al.*, 2021).

Menariknya, wacana di luar *domain* wisata juga menyumbang perspektif; Sakina Salmani dan Sarvesh Kulkarni menunjukkan bagaimana SVD++ dan CB pada MovieLens-100k bisa menurunkan RMSE ke 0,760. Meskipun studi ini berfokus pada film, konsep hibrida yang diusung menawarkan cetak biru bagi *domain* wisata terutama dalam penanganan data besar dan *heterogen* (Salmani & Kulkarni, 2021).

Jika ditarik benang merah, keseluruhan 15 studi tersebut menggambarkan evolusi progresif: dari eksploitasi model bahasa umum ke model *domain*-spesifik; dari *chatbot* berbasis aturan ke *chatbot* berdaya *deep learning*; dan dari rekomendasi *mono-algoritme* ke pendekatan hibrida yang memadukan data eksplisit, konten, demografi, hingga sentimen aspek. Tren terkini cenderung mengarah pada integrasi multibahasa, *modul voice*, serta *deployment* ringan berbasis *micro-services* atau *edge-device* untuk

menekan latensi. Meski demikian, sejumlah celah masih terbuka mulai keterbatasan dataset berlabel sentimen, kebutuhan standardisasi evaluasi lintas-*domain*, hingga isu privasi ketika profil pengguna kian *granular*. Kesemua temuan ini memperkaya landasan konseptual, sekaligus menegaskan urgensi merancang sistem rekomendasi wisata hibrida berbasis *chatbot* yang responsif, kontekstual, dan adaptif terhadap preferensi dinamis wisatawan modern.

Eksplorasi lebih mendalam atas kelima belas studi tersebut mengungkap pola metodologis yang saling melengkapi sekaligus memperlihatkan ruang inovasi baru. Secara teknis, tiga rumpun besar pengenalan bahasa *domain*, antarmuka percakapan, dan algoritma rekomendasi saling beririsan dalam beberapa studi, tetapi jarang terintegrasi penuh ke dalam satu prototipe *end-to-end*. Misalnya, Dhomas Hatta Fudholi *et al.* telah membuktikan kapasitas BERT untuk mengekstrak entitas wisata (Fudholi *et al.*, 2022), sementara Choirul Huda *et al.* mengilustrasikan bagaimana ABSA memperkaya penilaian kualitas destinasi (Huda *et al.*, 2024). Namun, belum ada riset yang secara sistematis menggabungkan BERT-NER dan ABSA dalam satu pipeline *chatbot* untuk mempersonalisasi rekomendasi berbasis sentimen *real-time*. Kesenjangan ini menawarkan peluang eksplorasi sistem terpadu yang mampu “mendengar” emosi wisatawan sekaligus “mengerti” konteks destinasi.

Dari sisi dataset, variasi skala dan *heterogenitas* tampak mencolok. Studi Zein Hanni Pradana *et al.* bekerja dengan 21 kategori FAQ yang relatif ringkas (Pradana *et al.*, 2022), sedangkan Saluja *et al.* memanfaatkan lebih dari 130 000 pasangan tanya-jawab (Saluja *et al.*, 2021). Ketimpangan ini berpengaruh langsung pada generalisasi model. TourBERT (Arefieva & Egger, 2021) menunjukkan bahwa *pre-training domain*-spesifik pada korpus jutaan dokumen drastis meningkatkan akurasi, tetapi pendekatan serupa belum diadopsi di banyak studi berbasis *chatbot*. Masih terbuka

pertanyaan: berapa volume data minimal agar model bahasa wisata bisa menyaingi TourBERT tanpa biaya komputasi ekstrem? Studi few-shot atau instruction-tuning berpotensi menjawab dilema tersebut sambil menekan jejak karbon pelatihan model.

Mengenai metrik evaluasi, hampir semua penelitian mengandalkan akurasi, *precision*, *recall*, *F1-score*, atau RMSE/MAE. Namun, hanya segelintir yang memasukkan metrik pengalaman pengguna seperti NASA-TLX (Sperli, 2021) atau loyalitas aplikasi (Iraola *et al.*, 2023). Padahal, efektivitas rekomendasi wisata tidak semata-mata diukur melalui skor prediksi, tetapi juga persepsi kegunaan, kepercayaan, dan kenyamanan. Kombinasi metrik kuantitatif dan kualitatif misalnya *long-term engagement*, *time-to-first-satisfactory-answer*, atau *conversion rate* pemesanan masih langka. Hal ini menegaskan perlunya kerangka evaluasi holistik yang mengaitkan performa algoritma dengan indikator keberhasilan bisnis pariwisata.

Di ranah algoritme, tren menuju hibridisasi tampak dominan. Huda *et al.* memadukan UBCF, CBCF, DF, dan ABSA (Huda *et al.*, 2024); Akbar *et al.* menimbang CB dan CF via *weighted parallel hybrid* (Akbar *et al.*, 2021); sementara Salmani & Kulkarni menunjukkan superioritas SVD++ di *domain* film (Salmani & Kulkarni, 2021). Meskipun berbeda *domain*, pola yang sama yakni menggabungkan beberapa sumber sinyal telah terbukti mengurangi *sparsity*. Tantangan berikutnya ialah menentukan bobot hibrida secara adaptif. Sebagian besar studi masih memakai skema bobot statis; riset mendatang dapat mengeksplor *multi-armed bandit* atau *meta-learning* untuk menyesuaikan bobot secara dinamis mengikuti perilaku pengguna.

Bahasan keamanan dan etika belum menjadi fokus utama, meski data yang dipakai kerap mencakup profil demografi (Huda *et al.*, 2024) dan lokasi *real-time* (Nguyen *et al.*, 2023). Implementasi kebijakan privasi berbasis *federated learning* atau

differential privacy hampir tidak tersentuh. Begitu pula isu bias misalnya bias lokasi populer yang mendominasi hasil rekomendasi masih sedikit dibedah selain sekadar mengakui keterbatasan dataset (Fudholi *et al.*, 2022; Saputra *et al.*, 2024). Dengan regulasi data global kian ketat, agenda penelitian seharusnya memasukkan audit bias dan penanganan *fair-recommendation* secara eksplisit.

Persilangan teknologi fisik dan digital juga menonjol: Puspa Miladin Safitri *et al.* mengintegrasikan *chatbot* ke Telegram (Safitri *et al.*, 2023), sedangkan aplikasi *IoT* pada *smart tourism* seperti sensor lokasi atau *smart lock* hotel belum banyak diadopsi di lini rekomendasi. Integrasi *IoT* dapat memperkaya konteks pergerakan wisatawan secara *real-time*, memungkinkan rekomendasi proaktif berbasis kedekatan geospasial dan kondisi keramaian. Akhirnya, lintasan riset menunjukkan evolusi bertahap ke arah sistem multimodal. “CHAT-Bot” oleh Casillo *et al.* sudah memadukan teks dan media (Casillo *et al.*, 2020), tetapi hanya TourBERT yang mengeksplor klaster visual secara mendalam (Arefieva & Egger, 2021). Dengan maraknya *vision-language models* (VLM) pasca-2023, penelitian wisata selanjutnya berpeluang mencangkok VLM untuk mengenali landmark secara visual dan menawarkan rekomendasi berbasis citra serta teks sekaligus.

2.21 Kesimpulan Tinjauan Studi

Penelitian terdahulu secara umum telah memberikan kontribusi signifikan dalam pengembangan sistem berbasis AI untuk sektor pariwisata, termasuk penerapan *chatbot* dan sistem rekomendasi. Sebagian besar penelitian mengadopsi teknologi algoritma *machine learning* serta kerangka kerja seperti RASA untuk menciptakan solusi interaktif yang mendukung pengalaman wisatawan. Fokus penelitian meliputi pengenalan entitas, rekomendasi tempat wisata, dan penyediaan informasi berbasis konteks. Meskipun demikian, studi-studi tersebut mengakui beberapa keterbatasan,

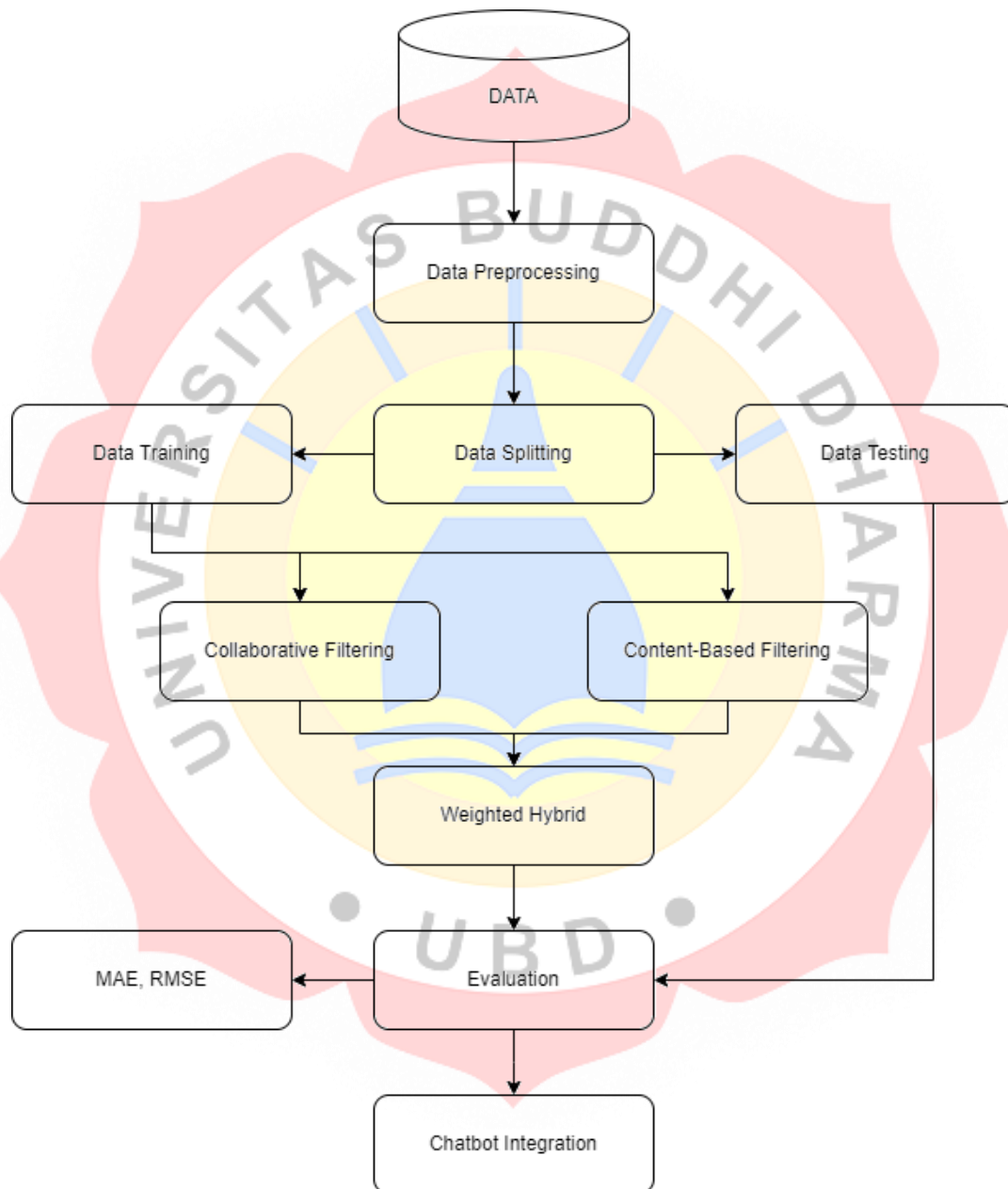
seperti cakupan data yang terbatas, keterbatasan multibahasa, dan kurangnya personalisasi berdasarkan preferensi pengguna. Dalam pengembangannya, teknologi seperti *hybrid filtering*, *deep learning*, dan algoritma *sequence-to-sequence* telah menunjukkan hasil yang menjanjikan dengan akurasi tinggi dalam prediksi dan rekomendasi.

Saat ini, posisi penelitian berada pada titik untuk menggabungkan berbagai pendekatan yang telah terbukti efektif dalam meningkatkan pengalaman wisatawan. Dengan memanfaatkan metodologi dan teknologi dari penelitian sebelumnya, pengembangan sistem rekomendasi pariwisata berbasis AI dapat lebih terintegrasi dan komprehensif. Penelitian ini berupaya mengatasi keterbatasan sebelumnya, seperti keterbatasan personalisasi yang lebih mendalam, dengan fokus pada peningkatan interaksi pengguna melalui kombinasi teknologi kecerdasan buatan, *chatbots*, *hybird filtering* yang lebih adaptif. Selain itu, penelitian ini bertujuan untuk memberikan solusi yang lebih efisien dan ramah pengguna guna mendukung digitalisasi industri pariwisata di Indonesia.

BAB III

METODE PENELITIAN

3.1 Kerangka Pemikiran



Gambar 3.1 Kerangka Pemikiran

Gambar 3.1 menunjukkan kerangka penelitian yang menggambarkan tahapan utama dalam pembangunan sistem rekomendasi destinasi wisata berbasis *hybrid filtering* yang terintegrasi dengan aplikasi *chatbot*. Kerangka ini dirancang untuk memberikan gambaran alur proses secara sistematis, dimulai dari tahap pengolahan data hingga integrasi ke dalam sistem *chatbot*. Tahap awal dimulai dari pengumpulan data, yang mencakup data destinasi wisata dan interaksi pengguna. Seluruh data yang diperoleh akan melalui proses data preprocessing, yaitu tahap pembersihan, normalisasi, pengisian nilai yang hilang, dan transformasi fitur agar data siap digunakan untuk proses pelatihan model. Setelah data diproses, dilakukan tahap data splitting, yaitu pemisahan data menjadi data pelatihan (*training*) dan data pengujian (*testing*). Data pelatihan akan digunakan untuk membangun dua model rekomendasi utama, yaitu: *Collaborative Filtering* (CF): Menggunakan interaksi pengguna (seperti *rating* atau preferensi eksplisit) untuk menemukan kemiripan antar pengguna maupun antar *item*. *Content-Based Filtering* (CBF): Mengandalkan deskripsi dan fitur dari destinasi wisata, seperti kategori, lokasi, dan atribut lainnya, untuk merekomendasikan *item* yang serupa dengan yang disukai pengguna sebelumnya. Hasil dari kedua pendekatan ini kemudian digabungkan menggunakan teknik *Weighted Hybrid*, yaitu metode yang mengombinasikan skor rekomendasi dari CF dan CBF berdasarkan bobot tertentu, untuk menghasilkan rekomendasi yang lebih akurat dan relevan. Model hybrid yang telah dibentuk kemudian dievaluasi pada tahap *Evaluation* menggunakan metrik kuantitatif seperti MAE dan RMSE. Metrik ini digunakan untuk mengukur akurasi prediksi sistem terhadap data pengujian. Tahap akhir dari kerangka ini adalah *Chatbot Integration*, yaitu proses integrasi sistem rekomendasi ke dalam antarmuka *chatbot* berbasis *framework* RASA. *Chatbot* ini memungkinkan pengguna berinteraksi secara natural dengan sistem, meminta rekomendasi destinasi wisata, serta menerima hasil rekomendasi secara langsung dalam bentuk percakapan. Secara keseluruhan, kerangka ini menggambarkan integrasi

antara pendekatan teknik rekomendasi modern dan teknologi *chatbot* sebagai solusi cerdas dalam membantu pengguna memilih destinasi wisata sesuai preferensi mereka.

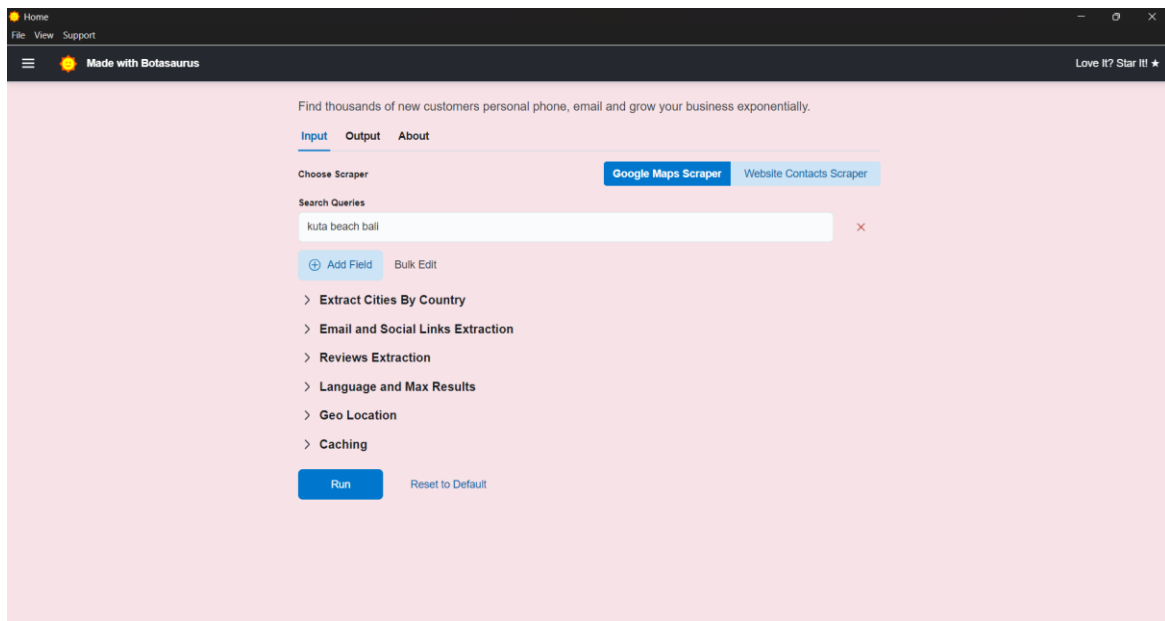
3.2 Data Collection

Pengumpulan data dalam penelitian ini dilakukan melalui proses *web scrapping* terhadap platform *Google Maps*, dengan menggunakan alat bantu berupa *Google Maps Extractor* yang tersedia sebagai proyek *open-source* di *GitHub*. Alat ini digunakan untuk mengotomatisasi proses ekstraksi informasi dari halaman tempat wisata, termasuk data umum dan ulasan pengguna.

Hasil *scrapping* berupa dua jenis data utama, yaitu:

- a. Data ringkasan tempat wisata, seperti nama, deskripsi, *rating* rata-rata, jumlah ulasan, dan informasi kontak, dan sebagainya
- b. Data ulasan terperinci, meliputi nama reviewer, teks ulasan, *rating* individu, tanggal publikasi, status sebagai local guide, dan sebagainya.

Proses *scrapping* berjalan melalui simulasi perilaku pengguna pada halaman *Google Maps* menggunakan *script* otomatis, dan data yang dihasilkan dalam bentuk dua *file CSV* yaitu, *wisata-di-indonesia-overview.csv*, dan *wisata-di-indonesia-detailed-reviews.csv*. Seluruh proses dilakukan dengan mempertimbangkan batasan teknis dan etika *scrapping*, seperti pembatasan frekuensi akses agar tidak melanggar kebijakan penggunaan *Google Maps*. Alat *scrapping* yang digunakan berasal dari proyek *open-source* *GitHub* dengan alamat sebagai berikut: <https://github.com/omkarcloud/google-maps-scraper>.



Gambar 3.2 Halaman input pada alat *google maps scrapper/extractor*

Place Id	Name	Description	Is Spendin...	Reviews	Rating	Competitors	Website	Phone	Can Claim	Owner Name	Owner Pr
ChiJQw23...	Pantai Kuta	Kawasan p...	False	41319	4.5		http://kutab...		False		
ChiJ36guP...	Pantai Nus...	Pantai tena...	False	1108	4.5				False		
ChiJZ5sY9...	Mandala S...	The Monke...	False	53279	4.5	Name: San...	https://mon...	0857-3700-...	False	Mandala S...	https://ww...
ChiJwRUJz...	Pantai Sanur	Perahu nel...	False	5837	4.5				False		
ChiJBf1Lg...	Pantai Sem...	Pantai berp...	False	4550	4.5				False		
ChiJq95xT...	Tanah Lot	Ticket Price...	False	97955	4.7	Name: Tan...	https://www...	(0361) 880...	False	Tanah Lot (...)	https://ww...
ChiJBa0pr...	Tanah Lot T...		False	376	4.8	Name: Tan...			True	Tanah Lot T...	
ChiJtezYD...	Sunset Bea...		False	316	4.7	Name: Tan...	https://www...	0812-3631-...	True	Sunset Bea...	
ChiJbyAH...	Pura Batu ...	Populer unt...	False	1420	4.6	Name: Pur...		(0361) 880...	False	Pura Batu ...	https://ww...
ChiJZMBY...	Tanah Lot		False	231	4.7				False		
ChiJuXDS...	Pura Pesim...		False	6	5	Name: Pur...			True	Pura Pesim...	
ChiJwIsn...	Pura Ulun ...	Dua buah b...	False	46179	4.6	Name: Pur...	https://ulun...	0821-4612-...	False	Pura Ulun ...	https://ww...

Gambar 3.3 Halaman hasil pada alat *google maps scrapper/extractor*

Gambar 3.2 menampilkan tampilan awal dari alat bantu *scrapping* Google Maps berbasis antarmuka grafis yang digunakan dalam penelitian ini. Pada halaman ini, pengguna dapat memasukkan kata kunci pencarian (*search query*) seperti “Kuta Beach Bali” untuk mencari lokasi wisata tertentu. Selain itu, tersedia berbagai opsi tambahan seperti ekstraksi ulasan

(*Reviews Extraction*), pengaturan bahasa dan jumlah hasil (*Language and Max Results*), serta opsi geolokasi dan *caching*. Proses *scrapping* dilakukan dengan menekan tombol *Run* setelah konfigurasi selesai. Gambar 3.3 menunjukkan hasil dari proses *scrapping Google Maps* berdasarkan *query* yang dimasukkan sebelumnya. Hasil ditampilkan dalam bentuk tabel yang mencakup berbagai atribut tempat wisata, seperti *Place Id*, *Name*, *Description*, *Reviews* (jumlah ulasan), *Rating*, *Website*, *Phone*, serta informasi tambahan seperti nama pemilik dan profil bisnis (jika tersedia). Data ini selanjutnya disimpan dalam format *CSV* dan digunakan sebagai dasar dalam pengembangan sistem rekomendasi tempat wisata.

3.3 Data Understanding

Penelitian ini menggunakan *dataset* primer yang berisi 1.048.576 ulasan wisata, 851.102 pengguna, 50 atraksi wisata, dan tingkat *sparsity* data sebesar 97.6%. *Dataset* ini ini merepresentasikan data publik yang tersedia di *Google Maps*, meliputi informasi umum tempat wisata serta ulasan-ulasan mendetail dari pengguna yang pernah berkunjung. Berikut merupakan struktur *dataset* dalam bentuk tabel.

Tabel 3.1 Data Ringkasan Tempat Wisata

Kolom	Deskripsi
<i>place_id</i>	ID unik untuk setiap tempat wisata.
<i>name</i>	Nama tempat wisata.
<i>description</i>	Deskripsi tempat wisata.
<i>is_spendig_on_ads</i>	Nilai <i>boolean</i> apakah wisata ini pernah melakukan iklan.
<i>reviews</i>	Jumlah total ulasan wisata tersebut.
<i>rating</i>	Jumlah rata-rata <i>rating</i> wisata tersebut (1-5).
<i>competitors</i>	Kompetitor tempat wisata.
<i>can_claim</i>	Nilai <i>boolean</i> bisa di klaim atau tidak.
<i>owner_name</i>	Nama pemilik tempat.
<i>owner_profile</i>	<i>Link</i> profile pemilik tempat.
<i>website</i>	Alamat <i>website</i> tempat wisata tersebut jika tersedia.
<i>phone</i>	Nomor telepon tempat wisata tersebut jika tersedia.
<i>featured_image</i>	<i>Link</i> gambar utama tempat wisata.
<i>main_category</i>	Kategori utama tempat tersebut.
<i>categories</i>	Beberapa kategori dari tempat tersebut.
<i>workday_timing</i>	Informasi jam tempat beroperasi.
<i>is_temporary_closed</i>	Nilai <i>boolean</i> apakah wisata ini ditutup sementara.
<i>closed_on</i>	Informasi hari tempat tidak beroperasi.
<i>address</i>	Informasi alamat lengkap dari tempat wisata.

<i>review_keywords</i>	Kata kunci yang diambil otomatis dari ulasan teks.
<i>links</i>	Link wisata tersebut.
<i>query</i>	Input query yang digunakan pada alat scrapping.

Tabel 3.2 Data Ulasan Detail

Kolom	Deskripsi
<i>place_id</i>	ID unik untuk setiap tempat wisata.
<i>place_name</i>	Nama tempat wisata
<i>review_id</i>	ID unik dari masing-masing ulasan.
<i>review_link</i>	Link menuju ulasan asli di <i>Google Maps</i>
<i>name</i>	Nama pengguna yang memberi ulasan
<i>reviewer_id</i>	ID unik <i>reviewer</i> .
<i>reviewer_profile</i>	Link ke profil pengguna.
<i>rating</i>	Rating yang diberikan pengguna.
<i>review_text</i>	Isi teks ulasan.
<i>published_at</i>	Timestamp publikasi ulasan.
<i>published_at_date</i>	Tanggal publikasi dalam format terpisah.
<i>response_from_owner_text</i>	Tanggapan dari pemilik (jika ada).
<i>Total_reviews_by_reviewer</i>	Total ulasan yang pernah diberikan <i>reviewer</i> .
<i>is_local_guide</i>	Menunjukkan apakah pengguna adalah <i>Local Guide</i> .
<i>review_translated_text</i>	Teks ulasan versi terjemahan (jika tersedia).
<i>experience_details</i>	Rincian pengalaman pengguna (jika tersedia).
<i>review_photos</i>	Link foto yang diunggah bersama ulasan (jika ada).

Tabel 3.1 menyajikan struktur data ringkasan dari tempat wisata yang diperoleh melalui proses *scrapping Google Maps*. Setiap entri mewakili satu tempat wisata dengan atribut yang mencakup identitas *place_id*, nama tempat, deskripsi singkat, *rating* rata-rata dari pengguna, jumlah rata-rata ulasan, serta informasi tambahan seperti situs web, nomor telepon, gambar utama, dan sebagainya. Data ini menjadi fondasi awal dalam proses pemodelan rekomendasi karena memuat informasi yang dapat digunakan untuk menilai popularitas dan karakteristik umum dari destinasi wisata yang tersedia. Tabel 3.2 menunjukkan struktur data ulasan pengguna yang dikumpulkan dari halaman *Google Maps*. Setiap baris dalam tabel ini mewakili satu ulasan individu yang mencakup identitas pengguna *reviewer_id*, *rating* yang diberikan, isi ulasan dalam bentuk teks, tanggal publikasi, serta informasi tambahan seperti status *local guide* dan tanggapan dari pemilik

tempat. Data ini menjadi komponen utama dalam sistem rekomendasi terutama dalam pendekatan *collaborative filtering*, dan *content-based filtering*.

3.4 Data Preprocessing

Data *preprocessing* merupakan tahap awal dalam proses pengolahan data yang bertujuan untuk menyiapkan data mentah menjadi data yang siap digunakan dalam proses pemodelan sistem rekomendasi. Tahapan ini sangat penting untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik, bersih, konsisten, serta sesuai dengan format yang dibutuhkan oleh algoritma pembelajaran mesin. Dalam penelitian ini, proses *preprocessing* dilakukan terhadap dua sumber data utama, yaitu dataset ulasan pengguna (*detailed reviews*) dan dataset informasi tempat wisata (*overview*).

3.4.1 Data Cleaning

Proses pembersihan data (*data cleaning*) dilakukan untuk memastikan kualitas data yang akan digunakan dalam pemodelan sistem rekomendasi berada dalam kondisi optimal, bebas dari gangguan seperti duplikasi, ketidakkonsistenan, maupun data tidak relevan. Tahapan *cleaning* diterapkan khususnya pada dataset ulasan pengguna (*detailed reviews*) yang diperoleh dari hasil *scrapping* platform *Google Maps*. Beberapa kolom yang tidak digunakan dalam proses analisis dan pemodelan dihapus dari dataset, antara lain: *review_link*, *reviewer_profile*, *response_from_owner_text*, *response_from_owner_ago*, *response_from_owner_date*, *total_number_of_reviews_by_reviewer*, *total_number_of_photos_by_reviewer*, *response_from_owner_translated_text*, *published_at*, *is_local_guide*, dan *review_photos*. Tabel 3.3 dan Tabel 3.4 Menunjukkan ulasan *dataset (detailed reviews)* field sebelum dan sesudah proses *data cleaning*.

Tabel 3.3 Informasi Dataset sebelum proses *cleaning*

<i>Column</i>	<i>Data Type</i>
<i>place_id</i>	<i>Object</i>

<i>place_name</i>	<i>Object</i>
<i>review_id</i>	<i>Object</i>
<i>review_link</i>	<i>Object</i>
<i>name</i>	<i>Object</i>
<i>reviewer_id</i>	<i>Object</i>
<i>reviewer_profile</i>	<i>Object</i>
<i>rating</i>	<i>Float64</i>
<i>review_text</i>	<i>Object</i>
<i>published_at</i>	<i>Object</i>
<i>published_at_date</i>	<i>Object</i>
<i>response_from_owner_text</i>	<i>Object</i>
<i>total_reviews_by_reviewer</i>	<i>Int64</i>
<i>is_local_guide</i>	<i>Float64</i>
<i>review_translated_text</i>	<i>Object</i>
<i>experience_details</i>	<i>Object</i>
<i>review_photos</i>	<i>Object</i>

Tabel 3.4 Informasi Dataset setelah proses *cleaning*

Column	Data Type
<i>place_id</i>	<i>Object</i>
<i>place_name</i>	<i>Object</i>
<i>review_id</i>	<i>Object</i>
<i>review_link</i>	<i>Object</i>
<i>name</i>	<i>Object</i>
<i>reviewer_id</i>	<i>Object</i>
<i>rating</i>	<i>Float64</i>
<i>review_text</i>	<i>Object</i>
<i>published_at_date</i>	<i>Object</i>
<i>review_translated_text</i>	<i>Object</i>
<i>experience_details</i>	<i>Object</i>

Seluruh baris duplikat yang memiliki nilai identik pada seluruh kolom dihapus menggunakan metode *drop_duplicates()* untuk menghindari bias dan redundansi pada model, namun tidak ditemukan data duplikat dalam dataset jadi jumlah baris tetap dan tidak berubah. Seluruh kolom bertipe teks (*object*) dikonversi ke huruf kecil (*lowercase*) untuk menjaga konsistensi representasi kata, kecuali kolom *place_id*, *review_id* dan *reviewer_id* yang dipertahankan dalam format aslinya guna menjamin

konsistensi dalam proses pemetaan entitas. Kolom-kolom teks utama seperti *review_text*, *review_translated_text*, dan *experience_details* yang memiliki nilai kosong diisi dengan *string* kosong (*nan*). Ini dilakukan agar tidak mengganggu proses transformasi teks saat *Content-Based Filtering* dibangun. Setelah proses *cleaning* selesai, dataset yang telah dibersihkan disimpan kembali dalam bentuk *CSV* dengan nama *cleaned_reviews.csv*. Dataset ini menjadi dasar dalam proses pemodelan *Collaborative Filtering* dan *Content-Based Filtering* yang dijelaskan pada tahap selanjutnya.

Hal yang sama juga dilakukan pada *dataset* tempat wisata (*overview.csv*). Dataset ini berisi informasi profil dari berbagai tempat wisata di Indonesia, yang nantinya akan digunakan sebagai sumber fitur dalam sistem rekomendasi berbasis konten (*Content-Based Filtering*). Beberapa *field* (kolom) seperti *is_spending_on_ads*, *query*, *competitors*, *can_claim*, *owner_name*, dan *owner_profile* diidentifikasi tidak memberikan kontribusi signifikan terhadap proses perhitungan kemiripan konten. Oleh karena itu, kolom-kolom tersebut dihapus dari dataset akhir. Tabel 3.5 dan Tabel 3.6 Menunjukkan *dataset* tempat wisata (*overview*) *field* sebelum dan sesudah proses *data cleaning*.

Tabel 3.5 Informasi *Dataset* wisata sebelum proses *cleaning*

Column	Data Type
<i>place_id</i>	<i>Object</i>
<i>name</i>	<i>Object</i>
<i>description</i>	<i>Object</i>
<i>is_spending_on_ads</i>	<i>Float64</i>
<i>reviews</i>	<i>Int64</i>
<i>rating</i>	<i>Float64</i>
<i>competitors</i>	<i>Object</i>
<i>can_claim</i>	<i>Float64</i>
<i>owner_name</i>	<i>Object</i>
<i>owner_profile</i>	<i>Object</i>
<i>website</i>	<i>Object</i>

<i>phone</i>	<i>Object</i>
<i>featured_image</i>	<i>Object</i>
<i>main_category</i>	<i>Object</i>
<i>categories</i>	<i>Object</i>
<i>workday_timing</i>	<i>Object</i>
<i>is_temporary_closed</i>	<i>Float64</i>
<i>closed_on</i>	<i>Object</i>
<i>address</i>	<i>Object</i>
<i>review_keywords</i>	<i>Object</i>
<i>links</i>	<i>Object</i>
<i>query</i>	<i>Object</i>

Tabel 3.6 Informasi *Dataset* wisata sesudah proses *cleaning*

Column	DataType
<i>place_id</i>	<i>Object</i>
<i>name</i>	<i>Object</i>
<i>description</i>	<i>Object</i>
<i>reviews</i>	<i>Int64</i>
<i>rating</i>	<i>Float64</i>
<i>website</i>	<i>Object</i>
<i>phone</i>	<i>Object</i>
<i>featured_image</i>	<i>Object</i>
<i>main_category</i>	<i>Object</i>
<i>categories</i>	<i>Object</i>
<i>workday_timing</i>	<i>Object</i>
<i>is_temporary_closed</i>	<i>Float64</i>
<i>closed_on</i>	<i>Object</i>
<i>address</i>	<i>Object</i>
<i>review_keywords</i>	<i>Object</i>
<i>links</i>	<i>Object</i>

Seluruh kolom yang berisi informasi teks seperti *description*, *main_category*, *categories*, dan *review_keywords* dinormalisasi menjadi huruf kecil (*lowercase*) dan dihapus karakter-karakter *non-alfabet* agar dapat diproses lebih baik dalam algoritma berbasis teks seperti *TF-IDF*. Hal ini juga dilakukan untuk menjaga konsistensi antar entri dan menghindari duplikasi fitur semantik akibat perbedaan kapitalisasi atau penulisan. Ditemukan bahwa nilai pada kolom *main_category* dan *categories*

memiliki label yang sangat bervariasi, bahkan untuk tempat wisata dengan jenis yang sama. Oleh karena itu, dilakukan proses manual *mapping* dan standarisasi label, misalnya menyamakan kategori seperti “Museum Sejarah”, “Museum Seni”, dan “Museum Budaya” ke dalam satu label umum yaitu “Museum”. Langkah ini dilakukan untuk mengurangi *noise semantik* dan memperkuat bobot kategori utama pada tiap tempat wisata.

Tabel 3.7 Mapping kategorikal

Kategori Asli	Kategori Standar (Mapping)
Museum, Museum Seni, Museum Seni Modern	Museum
Pantai, Pulau	Pantai
Gunung berapi	Gunung
Taman Kota, Taman Nasional, Taman Hiburan, Kolam Renang, Taman Kanak-kanak	Taman
Wahana Taman Hiburan, Pusat Rekreasi, Hotel Resor, Pasar Apung, Pujasera, Jalan Terkenal, Tujuan Wisata	Rekreasi
Cagar Alam, Dinas Kehutanan	Alam
Pura, Vihara, Tempat Ibadah	Religi
Teater Seni Pertunjukan, Akuarium	Pertunjukan
Titik Pemandangan, Keindahan	Pemandangan
Monumen, Candi, Keraton	Monumen
Tukang Pijat Refleksi, Salon Kecantikan, Salon Rambut, Terapis Pijat	Pijat & Spa
Bangunan Bersejarah	Sejarah

Beberapa baris yang memiliki konten deskripsi sangat singkat atau generik seperti “mulai buka tanggal sekian” tanpa menjelaskan karakteristik tempat, diidentifikasi sebagai entri tidak informatif. Entri semacam ini ditandai untuk dihapus atau diganti dengan deskripsi dari sumber eksternal seperti *Wikipedia*.

3.4.2 Data Filtering

Untuk meningkatkan kualitas data dalam pemodelan *Collaborative Filtering*, dilakukan penyaringan terhadap data interaksi pengguna dengan kriteria minimum jumlah *rating*. Pengguna yang memiliki kurang dari dua interaksi *rating* dihapus dari

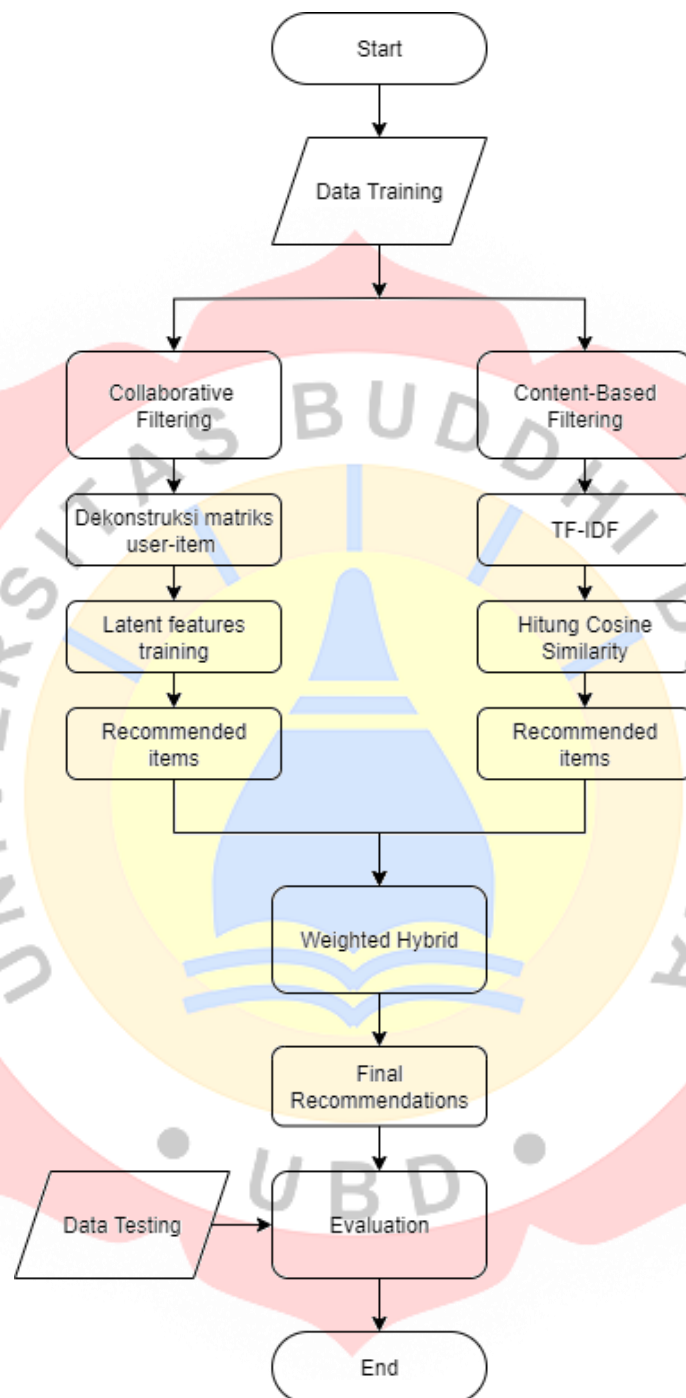
dataset ulasan. Hal ini bertujuan untuk mengurangi sparsitas pada *matriks user-item* dan memastikan bahwa model hanya dilatih pada pengguna yang memiliki preferensi yang cukup terwakili. Hasil dari penyaringan ini menunjukkan pengurangan jumlah pengguna dari 851.102 menjadi 52.482 pengguna dan pengurangan sebagian besar baris data ulasan dari 1.048.576 menjadi 434.420.

3.5 *Data Splitting*

Dalam penelitian ini, data yang telah diproses dibagi menjadi dua *subset*, yaitu *training set* dan *test set*, dengan proporsi yang umum digunakan masing-masing 80% dan 20% (Joseph, 2022). Pembagian ini dilakukan secara acak untuk memastikan bahwa model dapat belajar dari data yang representatif (*training set*) dan diuji pada data yang belum pernah dilihat sebelumnya (*test set*), sehingga kemampuan generalisasi model dapat dievaluasi secara objektif (Meng et al., 2020).

3.6 Model Development

3.6.1 Flowchart Model



Gambar 3.4 Flowchart Model

Gambar 3.4 menunjukkan alur pengembangan model, proses dimulai dengan *data training*, di mana data yang telah diproses sebelumnya digunakan untuk melatih

dua model utama, yaitu *Collaborative Filtering* (CF) dan *Content-Based Filtering* (CBF). Pada jalur *Collaborative Filtering*, model melakukan dekomposisi *matriks user-item* menggunakan metode *Matrix Factorization*, seperti *Singular Value Decomposition* (SVD++). Proses ini bertujuan untuk menemukan *latent features* (fitur tersembunyi) yang merepresentasikan hubungan antara pengguna dan *item* (Chanrueang et al., 2024). Hasil dari tahap ini adalah rekomendasi *item* yang relevan untuk pengguna berdasarkan pola interaksi dengan pengguna lain. Di jalur CBF, dilakukan proses *TF-IDF* untuk merepresentasikan atribut *item* (seperti kategori dan lokasi) dalam bentuk vektor. Proses ini diikuti oleh perhitungan *cosine similarity* untuk mengukur kesamaan antar *item* berdasarkan preferensi pengguna. Hasil akhirnya adalah rekomendasi *item* yang sesuai dengan karakteristik *item* yang telah dinilai positif oleh pengguna. Rekomendasi yang dihasilkan oleh kedua metode (CF dan CBF) digabungkan menggunakan pendekatan *Weighted Hybrid*, di mana bobot tertentu diberikan untuk setiap metode berdasarkan kontribusinya (Mehrabani et al., 2020). Hasil penggabungan ini menghasilkan rekomendasi akhir yang memadukan keunggulan kedua pendekatan.

3.6.2 *Collaborative Filtering*

Collaborative Filtering (CF) adalah pendekatan yang memanfaatkan hubungan antar pengguna dan *item* berdasarkan pola *rating* yang telah diberikan. Pada penelitian ini, CF diterapkan menggunakan teknik *Matrix Factorization* dengan *Singular Value Decomposition* (SVD++), yang merupakan salah satu metode terkemuka dalam sistem rekomendasi (Amri Akbar et al., 2021). Pada tahap ini, *dataset* utama masih diperlukan langkah penyaringan untuk memastikan hanya data yang relevan digunakan dalam model.

Collaborative Filtering menggunakan kolom *reviewer_id*, *places_id*, dan *rating* yang digunakan untuk membangun User-Item Based *Collaborative Filtering* yang menggunakan Matrix Factorization dengan Singular Value Decomposition. Tabel 3.8 menunjukkan hasil penyaringan.

Tabel 3.8 Tabel CF

<i>Coloumn</i>	<i>Non-Null Count</i>	<i>Dtype</i>
<i>reviewer_id</i>	52930 non-null	object
<i>places_id</i>	52930 non-null	object
<i>rating</i>	52930 non-null	int64

Proses pengembangan model *Collaborative Filtering* melibatkan langkah-langkah berikut:

1. *Dataset training* yang terdiri dari kolom *reviewer_id*, *places_id*, dan *rating* digunakan untuk membentuk *matriks user-item*. Elemen dalam *matriks* merepresentasikan *rating* yang diberikan pengguna terhadap *item* tertentu, sedangkan elemen kosong menunjukkan *rating* yang belum tersedia.
2. Dengan menggunakan SVD++, *matriks user-item R* didekomposisi menjadi tiga *matriks*.
3. Matriks hasil rekonstruksi digunakan untuk memprediksi *rating* yang belum diketahui, yang kemudian menjadi dasar rekomendasi.
4. Model CF diimplementasikan menggunakan *library Surprise* dengan parameter utama seperti jumlah faktor laten (*n_factors*), jumlah iterasi pelatihan (*n_epochs*), dan kecepatan pembelajaran (*learning_rate*).

3.6.3 Content-Based Filtering

Content-Based Filtering (CBF) adalah pendekatan sistem rekomendasi yang memberikan saran kepada pengguna berdasarkan kesamaan atribut dari *item* yang sebelumnya disukai. Dalam konteks penelitian ini, metode CBF digunakan untuk

merekomendasikan tempat wisata dengan memanfaatkan informasi dari kategori dan lokasi tempat wisata yang pernah dinilai positif oleh pengguna.

Model ini mengandalkan dua fitur utama dari setiap tempat wisata, yaitu:

- a. *categories*: yang mencerminkan jenis dan karakteristik tempat wisata.
- b. *location*: informasi lokasi geografis tempat wisata, seperti nama kota atau daerah.

Tabel 3.9 Tabel CBF

<i>Coloumn</i>	<i>Non-null Count</i>	<i>Dtype</i>
<i>place_id</i>	52930 non-null	object
<i>categories</i>	52930 non-null	object
<i>location</i>	52930 non-null	object

Adapun langkah-langkah pengembangan model *Content-Based Filtering*:

- a. Pembersihan dan Persiapan Data Data tempat wisata dibersihkan dari nilai kosong dan disiapkan dalam dua atribut teks utama: *categories* dan *location*. Atribut tersebut diproses menjadi bentuk teks yang siap untuk ekstraksi fitur.
- b. Ekstraksi Fitur dengan *TF-IDF* Setiap tempat wisata direpresentasikan sebagai vektor numerik menggunakan teknik *TF-IDF* (*Term Frequency–Inverse Document Frequency*), untuk masing-masing fitur kategori dan lokasi. Proses ini menghasilkan representasi fitur yang mencerminkan seberapa penting istilah tertentu dalam konteks tempat wisata.
- c. Pembobotan dan Penggabungan Fitur Vektor hasil ekstraksi dari kategori dan lokasi kemudian dikombinasikan menjadi satu *matriks* fitur akhir. Bobot yang lebih tinggi diberikan pada fitur kategori (0.7), sedangkan lokasi diberi bobot (0.3), untuk menyesuaikan dengan relevansi preferensi pengguna terhadap jenis tempat wisata.

- d. Pembangunan Profil Pengguna Profil setiap pengguna dibentuk berdasarkan tempat wisata yang diberi *rating* tinggi ($rating \geq 4.0$). Fitur dari tempat-tempat tersebut dirata-ratakan untuk menghasilkan representasi vektor yang merefleksikan minat pengguna secara keseluruhan.
- e. Perhitungan Kemiripan dan Rekomendasi Sistem menghitung kemiripan antara profil pengguna dan setiap tempat wisata menggunakan *Cosine Similarity*. Tempat-tempat dengan tingkat kemiripan tertinggi kemudian direkomendasikan sebagai destinasi yang relevan.
- f. Implementasi Sistem Sistem dibangun menggunakan bahasa pemrograman *Python*, dengan pustaka-pustaka utama seperti: *pandas* dan *numpy* untuk pengolahan data, *scikit-learn* untuk transformasi *TF-IDF* dan perhitungan *Cosine Similarity*, *joblib* untuk serialisasi dan penyimpanan model

3.6.4 Knowledge Based

Salah satu pendekatan yang digunakan dalam penelitian ini adalah sistem rekomendasi berbasis pengetahuan atau *Knowledge-Based Recommender System* (KBSRS). Sistem ini memberikan rekomendasi berdasarkan pengetahuan eksplisit mengenai preferensi pengguna dan karakteristik *item*, tanpa memerlukan riwayat interaksi sebelumnya. Dalam konteks penelitian ini, sistem KBSRS dirancang untuk merekomendasikan tempat wisata berdasarkan *intent* yang didefinisikan oleh pengguna, seperti kategori wisata, lokasi, dan nilai *rating* minimum. Sistem KBSRS bekerja dengan cara memfilter *item* berdasarkan atribut-atribut yang secara eksplisit ditentukan oleh pengguna atau sistem eksternal (misalnya *chatbot*). Berbeda dengan pendekatan berbasis kolaboratif yang bergantung pada pola interaksi historis, pendekatan berbasis pengetahuan cocok digunakan dalam kondisi *cold-start*, yaitu ketika data interaksi masih terbatas atau tidak tersedia. Pada sistem yang

dikembangkan, pengguna dapat memberikan input berupa deskripsi preferensi, misalnya “berikan rekomendasi wisata alam di bali”. Input ini kemudian diproses oleh *chatbot* atau modul pemroses *intent* untuk menghasilkan struktur informasi eksplisit dalam bentuk data JSON, seperti yang bisa dilihat pada Tabel 3.11 *intent* dan entitas yang di ekstrak dari user input “berikan rekomendasi wisata alam di bali”.

Tabel 3.10 Informasi *intent* dan entitas

<i>Intent</i>	<i>Entitas</i>
“category”	“alam”
“location”	“bali”

3.6.5 *Popularity-Based*

Metode *Popularity-Based Recommendation* merupakan pendekatan paling sederhana dalam sistem rekomendasi, di mana *item* yang paling populer secara keseluruhan akan direkomendasikan kepada semua pengguna, tanpa mempertimbangkan preferensi individu. Model ini sangat berguna untuk menangani kasus *cold-start* pada pengguna baru yang belum memiliki riwayat interaksi (*rating*), sehingga belum memungkinkan untuk dilakukan personalisasi.

Pada sistem ini, popularitas suatu tempat wisata dihitung berdasarkan jumlah *review* dan rata-rata *rating* yang diberikan oleh seluruh pengguna. Tempat wisata yang memiliki jumlah *rating* tinggi dan nilai rata-rata yang baik akan diprioritaskan dalam rekomendasi. Langkah-langkah pengembangan model ini meliputi:

- Pengumpulan Data Ulasan mengambil data *user_id*, *place_id*, dan *rating* dari dataset.
- Perhitungan Popularitas Data dikelompokkan berdasarkan *place_id* untuk menghitung rata-rata *rating* (*avg_rating*) dan jumlah ulasan (*count*).

- c. Penggabungan Metadata Hasil perhitungan digabung dengan metadata tempat wisata dari dataset *places*, termasuk *name*, *location*, dan *categories*.
- d. Pemeringkatan dan Seleksi Tempat wisata diurutkan berdasarkan *avg_rating* tertinggi dan diambil *Top-N* sebagai rekomendasi.

3.6.6 Hybrid Approach

Pendekatan *Hybrid Filtering* dalam penelitian ini dikembangkan untuk menggabungkan kekuatan dua metode rekomendasi utama, yaitu *Collaborative Filtering* (CF) dan *Content-Based Filtering* (CBF), dalam satu sistem terpadu. Tujuannya adalah untuk meningkatkan akurasi dan relevansi rekomendasi, serta mengatasi tantangan *sparsity* dan *cold-start* yang umum terjadi pada sistem rekomendasi.

Sistem *hybrid* ini menggabungkan skor prediksi dari CF dan CBF melalui pendekatan *weighted hybrid* (late fusion), di mana skor dari masing-masing model dikombinasikan menggunakan bobot tertentu untuk menghasilkan peringkat akhir. Selain penggabungan skor, sistem ini juga dilengkapi dengan switching mechanism yang berfungsi untuk menentukan metode yang paling sesuai berdasarkan data yang tersedia dari pengguna:

- a. Jika pengguna memiliki ≥ 5 interaksi (*rating*): sistem menggunakan *Hybrid CF + CBF*.
- b. Jika pengguna memiliki 2–4 interaksi: sistem hanya menggunakan CBF
- c. Jika pengguna tidak memiliki interaksi atau sangat minim data: sistem menggunakan pendekatan *fallback* berbasis *Popularity-Based Recommendation*.

- d. Jika pengguna menyertakan preferensi eksplisit seperti lokasi, dan kategori maka sistem menerapkan *Knowledge-Based Filtering* sebagai filter akhir terhadap hasil rekomendasi dari model utama.

3.7 Metode Evaluasi

3.7.1 *Mean Absolute Error*

Mean Absolute Error (MAE) adalah metrik evaluasi yang digunakan untuk mengukur rata-rata kesalahan absolut antara nilai aktual dan nilai prediksi. MAE menghitung seberapa jauh prediksi model menyimpang dari nilai aktual tanpa memperhitungkan arah kesalahan (positif atau negatif). Metrik ini mudah diinterpretasikan karena nilainya langsung merepresentasikan rata-rata besarnya kesalahan dalam satuan *rating*. MAE sangat cocok untuk mengevaluasi sistem rekomendasi karena memberikan gambaran sederhana tentang keakuratan prediksi model.

3.7.2 *Root Mean Squared Error*

Root Mean Squared Error (RMSE) adalah metrik evaluasi yang mengukur akar rata-rata kuadrat kesalahan antara nilai aktual dan prediksi. RMSE memberikan bobot lebih besar pada kesalahan yang lebih besar, sehingga lebih sensitif terhadap outlier dibandingkan MAE. Hal ini membuat RMSE berguna untuk mengidentifikasi prediksi yang sangat tidak akurat dalam model. Nilai RMSE selalu lebih besar atau sama dengan MAE, dan semakin kecil nilai RMSE, semakin baik performa model dalam memprediksi nilai *rating*. Metrik ini sering digunakan dalam sistem rekomendasi untuk mengevaluasi kemampuan model menangani variasi yang besar dalam data.

3.7.3 *Precision, Recall, F1-Score*

Metrik evaluasi *Precision@K*, *Recall@K*, dan *F1-Score@K* untuk mengukur relevansi rekomendasi yang dihasilkan oleh sistem, khususnya pada pendekatan *Content-Based Filtering* (CBF), *Knowledge-Based Recommender* (KBR), dan *Popularity-Based Recommender*. *Precision@K* digunakan untuk mengukur sejauh mana daftar rekomendasi berisi *item* yang benar-benar relevan bagi pengguna. Nilai *Precision@K* dihitung sebagai proporsi *item* relevan dibandingkan dengan seluruh *item* yang direkomendasikan pada daftar *top-K*. *Precision* yang tinggi menunjukkan bahwa sebagian besar *item* yang ditampilkan memiliki tingkat kesesuaian tinggi terhadap preferensi pengguna, sehingga relevan pada skenario di mana jumlah *item* yang ditampilkan terbatas. *Recall@K* mengukur proporsi *item* relevan yang berhasil ditangkap oleh sistem dari seluruh *item* relevan yang tersedia. Nilai *Recall@K* yang tinggi menandakan cakupan rekomendasi yang luas, di mana sistem mampu menjaring sebagian besar *item* yang sesuai dengan preferensi pengguna, meskipun tidak semua *item* relevan berada di urutan teratas. *F1-Score@K* merupakan *harmonic mean* dari *Precision@K* dan *Recall@K*, yang digunakan untuk menyeimbangkan antara tingkat presisi dan cakupan. Metrik ini penting ketika kedua aspek tersebut sama-sama memiliki prioritas dalam evaluasi, sehingga *F1-Score* memberikan gambaran kinerja sistem secara menyeluruh dalam menghasilkan rekomendasi yang akurat sekaligus mencakup sebagian besar pilihan relevan.

3.8 Lingkungan Pengembangan

Penelitian ini dilakukan dengan dukungan perangkat keras dan lunak yang memadai untuk menunjang proses pengolahan data, pelatihan model sistem rekomendasi, serta evaluasi performa sistem. Spesifikasi ini disusun untuk menjamin efisiensi dan keandalan dalam eksekusi program, khususnya ketika menangani dataset berukuran besar, melakukan

ekstraksi fitur berbasis teks (*TF-IDF*), dan pelatihan algoritma seperti *Content-Based Filtering* (CBF) maupun *Collaborative Filtering* (CF).

3.8.1 Spesifikasi *Hardware*

Penelitian ini dijalankan menggunakan komputer pribadi (PC) dengan spesifikasi perangkat keras sebagaimana ditampilkan dalam Tabel 3.10. Spesifikasi ini mencakup prosesor berperforma tinggi dan memori yang cukup besar untuk menangani beban komputasi dari pemrosesan *dataset* dan pelatihan model *machine learning*.

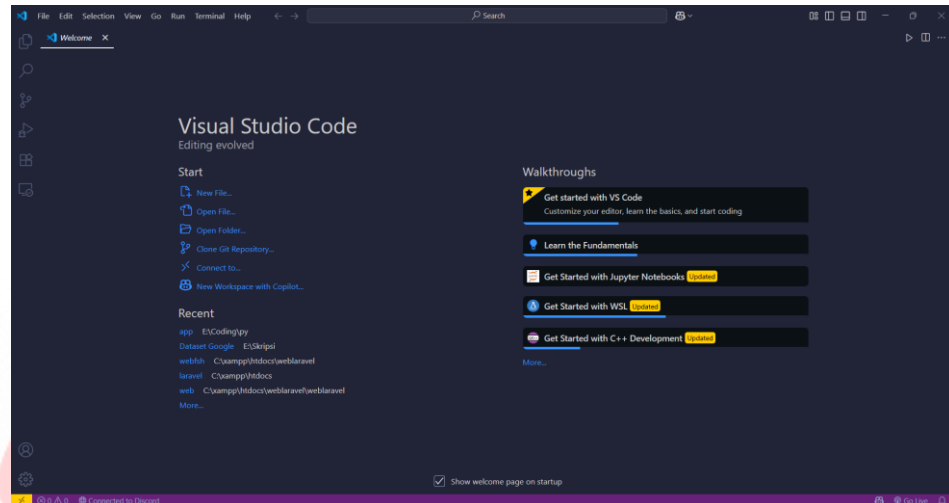
Tabel 3.11 Spesifikasi *Hardware*

Komponen	Spesifikasi
<i>Processor (CPU)</i>	<i>Intel® Core™ i5-11400H 2.3GHz</i>
Memori (<i>RAM</i>)	<i>16 GB DDR4 3200 MHz</i>
Penyimpanan (<i>Storage</i>)	<i>1 TB NVMe SSD (Solid State Drive)</i>
Kartu Grafis (<i>GPU</i>)	<i>NVIDIA GeForce RTX 3050 4GB GDDR5</i>
Sistem Operasi	<i>Windows 11 Pro 64-bit</i>
Monitor	<i>15.6" FHD IPS Display</i>

3.8.2 Spesifikasi *Software*

Perangkat lunak yang digunakan dalam penelitian ini terdiri dari sistem operasi *Windows 11*, lingkungan pemrograman *Python 3.9*, serta sejumlah pustaka (*libraries*) seperti *pandas*, *numpy*, *scikit-learn*, dan *Sastrawi* yang digunakan untuk pemrosesan data dan pembentukan model sistem rekomendasi berbasis *Content-Based Filtering*. Penelitian ini juga menggunakan *Visual Studio Code* untuk proses pemodelan dan pelatihan yang dapat dilihat pada Gambar 3.5. Selain itu, penelitian ini juga

menggunakan alat bantu *Google Maps Extractor* sebagai *tools scrapping* untuk mengumpulkan data ulasan dan metadata tempat wisata dari platform *Google Maps*.



Gambar 3.5 Visual Studio Code

3.9 Chatbot Development

Pengembangan *chatbot* dalam penelitian ini bertujuan untuk menyediakan antarmuka percakapan interaktif yang memungkinkan pengguna memperoleh rekomendasi tempat wisata secara kontekstual dan personal. Sistem ini diimplementasikan menggunakan *RASA Framework*, yang dipilih karena kemampuannya dalam menangani pemrosesan bahasa alami (*Natural Language Understanding*) berbasis pembelajaran mesin serta mendukung integrasi aksi khusus (*custom actions*) dengan sistem backend eksternal (T. T. Nguyen *et al.*, 2021).

3.9.1 Arsitektur dan Teknologi

Arsitektur *chatbot* terdiri dari tiga komponen utama:

- a. *Natural Language Understanding (NLU)*: Menangani deteksi *intent* dan ekstraksi entitas seperti lokasi, kategori wisata.

- b. *Dialogue Management*: Mengatur alur percakapan berdasarkan *intent* dan rules yang telah didefinisikan.
- c. *Custom Actions (Backend Integration)*: Menyambungkan *chatbot* ke API sistem rekomendasi wisata yang dibangun secara modular menggunakan *FastAPI*.

Model NLU RASA menggunakan pendekatan pipeline berbasis *DIETClassifier*, serta beberapa komponen tambahan seperti *WhitespaceTokenizer*, *RegexFeaturizer*, dan *EntitySynonymMapper*. Dalam pipeline ini, *RASA* dapat memahami berbagai variasi permintaan pengguna dengan tingkat akurasi yang cukup tinggi. *RASA* juga mendukung penggunaan *model BERT* secara terintegrasi dalam komponennya, namun dalam implementasi ini digunakan model default pipeline berbasis *count vectors* dan *embeddings* lokal yang lebih ringan namun tetap kontekstual (H. T. Nguyen *et al.*, 2023).

3.9.2 Deteksi *Intent* dan Entitas

Intent utama yang ditangani oleh *chatbot* meliputi:

- a. *minta_rekomendasi*: untuk permintaan saran tempat wisata.
- b. *tanya_info_tempat*: untuk menanyakan deskripsi, alamat, jam buka, atau hari tutup tempat wisata.
- c. *tolak_kompetensi*: untuk menolak permintaan di luar lingkup *chatbot*.

Setiap *intent* didukung oleh contoh ekspresi yang dilatih di file *nlu.yml*, serta entitas seperti:

- a. *lokasi*: kota atau daerah tujuan wisata.

- b. kategori: jenis wisata seperti pantai, gunung, taman, dll.
- c. *place_name, info_type*: untuk informasi spesifik tempat wisata.

3.9.3 Aksi Kustom dan Integrasi API

Untuk menjalankan sistem rekomendasi, *chatbot* memanggil *custom action* *action_rekomendasi*, yang mengirim permintaan HTTP POST ke *endpoint /recommend* pada *FastAPI*. *Payload* JSON mencakup *user_id* dan objek *intent* yang disusun dari slot pengguna. API kemudian menentukan model rekomendasi terbaik menggunakan fungsi *choose_best_model()* berdasarkan jumlah interaksi pengguna, dengan kemungkinan model:

- a. *Hybrid Filtering* (CF + CBF),
- b. *CBF-only* (jika data interaksi belum cukup),
- c. *Popularity-Based* (untuk cold-start),
- d. Disaring menggunakan *Knowledge-Based Recommender* berdasarkan *intent*.

Hasil dari API dikembalikan ke *chatbot* dan disampaikan ke pengguna dalam bentuk daftar rekomendasi yang telah diformat secara interaktif.

3.9.4 Rules dan Story

Pengaturan alur percakapan dalam *chatbot* dilakukan dengan memanfaatkan dua pendekatan utama yang disediakan oleh RASA, yaitu *rules* dan *stories*. *Rules* digunakan untuk mendefinisikan respon statis terhadap *intent* tertentu yang memiliki pola interaksi sederhana dan konsisten, seperti ketika pengguna meminta rekomendasi atau menanyakan informasi tempat. Misalnya, jika *chatbot* mendeteksi *intent* *minta_rekomendasi*, maka sistem akan langsung menjalankan aksi *action_rekomendasi* tanpa mempertimbangkan konteks percakapan sebelumnya. Di sisi lain, *stories* digunakan untuk mengatur alur dialog yang lebih kompleks atau kontekstual, terutama ketika *chatbot* harus mengingat tahapan interaksi sebelumnya.

Stories memungkinkan model memahami urutan percakapan dan memberikan respons yang lebih alami dan berurutan berdasarkan riwayat dialog.

Model RASA kemudian dilatih dengan menggunakan gabungan *RulePolicy* dan *TEDPolicy*. *RulePolicy* bertanggung jawab untuk menjalankan alur yang telah ditentukan secara eksplisit dalam *rules*, sedangkan *TEDPolicy* memungkinkan *chatbot* belajar pola percakapan dari data *training* dan menangani skenario yang tidak tercakup dalam *rules*. Kombinasi kedua strategi ini memastikan bahwa *chatbot* dapat memberikan respons yang stabil, adaptif, dan dapat dijelaskan, baik dalam kondisi dialog langsung maupun bercabang.

