

**PENGEMBANGAN SISTEM REKOMENDASI ADAPTIF
REAL-TIME UNTUK DESTINASI WISATA: OPTIMALISASI
KEBERAGAMAN DAN CAKUPAN GUNA MENDORONG
PEMERATAAN DISTRIBUSI PARIWISATA**

PROPOSAL TESIS

**Karya tulis sebagai salah satu syarat
kelulusan MK IF5002 Penyusunan Proposal**

**Oleh
M EGYPT PRATAMA
NIM: 22523311
(Program Studi Magister Informatika)**



**INSTITUT TEKNOLOGI BANDUNG
Maret 2025**

**PENGEMBANGAN SISTEM REKOMENDASI ADAPTIF
REAL-TIME UNTUK DESTINASI WISATA: OPTIMALISASI
KEBERAGAMAN DAN CAKUPAN GUNA MENDORONG
PEMERATAAN DISTRIBUSI PARIWISATA**

Oleh
M Egypt Pratama
NIM: 23523311
(Program Studi Magister Informatika)

Institut Teknologi Bandung

Menyetujui
Calon Tim Pembimbing

Tanggal 25 Maret 2025

Calon Pembimbing I

Dr. Ir. Arry Akhmad Arman, M.T.

DAFTAR ISI

HALAMAN PENGESAHAN	i
DAFTAR ISI	ii
DAFTAR LAMPIRAN.....	iv
DAFTAR GAMBAR DAN ILUSTRASI.....	v
DAFTAR TABEL.....	vi
DAFTAR SINGKATAN DAN LAMBANG	vii
BAB I. Pendahuluan	1
I.1 Latar Belakang	1
I.2 Rumusan Masalah	5
I.3 Tujuan Penelitian	5
I.4 Ruang Lingkup Masalah	6
I.5 Kebaruan dan Kontribusi	6
I.6 Sistematika Penulisan	7
BAB II. Tinjauan Pustaka	8
II.1 Sistem rekomendasi	8
II.2 Aplikasi dan implementasi.....	9
II.3 Klasifikasi Sistem Rekomendasi.....	10
II.3.1 <i>Collaborative filtering</i>	10
II.3.2 <i>Content-based filtering</i>	11
II.3.3 <i>Location-based</i>	11
II.3.4 <i>Context-aware</i>	11
II.3.5 <i>Hybrid</i>	12
II.4 Metrik Evaluasi	12
II.4.1 Metrik Akurasi Prediksi	12
II.4.2 Metrik Akurasi Klasifikasi.....	13
II.4.3 Metrik Akurasi Reranking	14
II.4.4 Metrik Berbasis Peringkat.....	16
II.4.5 Metrik Lainnya.....	17
II.5 <i>Machine Learning</i>	20
II.5.1 <i>Supervised Learning</i>	20
II.5.2 <i>Unsupervised Learning</i>	22
II.6 Basis Data	23
II.6.1 Jenis Basis Data	24
II.6.2 Jenis –Jenis Data	24
II.7 <i>Incremental Learning</i>	25
II.8 <i>Data Streaming Process</i>	26
II.9 Penelitian terkait	26
BAB III. Metodologi Penelitian.....	30
III.1 Metode Penelitian DSRM.....	30
III.2 Identifikasi Masalah dan Motivasi	31
III.3 Menentukan Solusi dari Tujuan.....	32
III.4 Perancangan dan Pengembangan	32
III.5 Demonstrasi	33
III.6 Evaluasi	33
III.7 Komunikasi.....	33

BAB IV. Analisis Dan Desain Awal.....	34
IV.1 Analisis Kebutuhan Sistem.....	34
IV.1.1 Kebutuhan Fungsional	34
IV.1.2 Kebutuhan Non-Fungsional	34
IV.2 Model Konseptual Sistem.....	35
IV.3 Rencana Jadwal Penelitian	37
DAFTAR PUSTAKA	39

DAFTAR LAMPIRAN

DAFTAR GAMBAR DAN ILUSTRASI

Gambar III.1 Kerangka Kerja DSRM berdasarkan Peffers dkk (2007).....	30
Gambar III.2 Kerangka Penelitian	31
Gambar IV.1 Arsitektur sistem yang akan dikembangkan	35

DAFTAR TABEL

Table II.1 Klasifikasi kemungkinan hasil rekomendasi.....	12
Table IV.1 Rencana Jadwal Penelitaian.....	38

DAFTAR SINGKATAN DAN LAMBANG

SINGKATAN	Nama	Pemakaian pertama kali pada halaman
API	Application Programming Interface	Hal. 25
DSRM	Design Science Research Methodology	Hal. 23
MAE	Mean Absolute Error	Hal. 2
ML	Machine Learning	Hal. 3
NDCG	Normalized Discounted Cumulative Gain	Hal. 2
NoSQL	Not Only SQL	Hal. 17
ODTW	Objek Daya Tarik Wisata	Hal. 29
RBP	Rank Biased Precision	Hal. 13
RDBMS	Relational Database Management System	Hal. 18
RMSE	Root Mean Squared Error	Hal. 2
RS	Recommender Systems	Hal. 7
SVD	Singular Value Decomposition	Hal. 8

LAMBANG	Deskripsi	Halaman
α	Sudut atau parameter tertentu dalam persamaan (tergantung konteks)	Hal. 21
b	Basis logaritma untuk perhitungan DCG/NDCG	Hal. 19
f_n	False Negative (item relevan yang tidak direkomendasikan)	Hal. 16
f_p	False Positive (item tidak relevan yang direkomendasikan)	Hal. 16
N	Jumlah total item atau data (konteks tergantung formula)	Hal. 13
p	Parameter discount untuk Rank Biased Precision (RBP)	Hal. 19

LAMBANG	Deskripsi	Halaman
Q	Jumlah rekomendasi yang dihasilkan (tergantung konteks di naskah)	Hal. 20
$r_{u,i}$	Rating sesungguhnya dari pengguna	Hal. 14
$\tilde{r}_{u,i}$	Rating yang diprediksi oleh sistem	Hal. 14
t_n	True Negative (item tidak relevan dan tidak direkomendasikan)	Hal. 16
t_p	True Positive (item relevan yang direkomendasikan)	Hal. 16

BAB I. Pendahuluan

Bab ini menjelaskan dasar-dasar yang membantu dalam melakukan penelitian ini. Dasar pemikiran didasarkan pada fenomena masalah yang muncul dan dibahas dalam penelitian setelah membaca sumber referensi yang tersedia.

I.1 Latar Belakang

Pariwisata adalah salah satu sektor utama yang menggerakkan perekonomian global. Berdasarkan laporan World Tourism Organization (2024), industri ini menyumbang sekitar USD 3,4 triliun terhadap *Tourism Direct Gross Domestic Product* (TDGDP), atau sekitar 3% dari PDB dunia. Di era digital, kebiasaan wisatawan berubah drastis. Teknologi informasi tidak hanya mempermudah perencanaan perjalanan, tetapi juga memengaruhi keputusan wisatawan sebelum, selama, dan setelah perjalanan (Pencarelli, 2020). Sejalan dengan itu, lebih dari 50% perusahaan perjalanan kini mengutamakan sistem rekomendasi berbasis kecerdasan buatan untuk memberikan rekomendasi destinasi dan aktivitas yang sesuai dengan preferensi individu (Statista Research Department, 2025). Pemahaman mendalam mengenai preferensi pengguna baik yang bersifat eksplisit (seperti rating atau ulasan sebelumnya) maupun implisit (seperti riwayat pencarian atau interaksi), serta yang bersifat statis (misalnya, demografi atau tipe wisatawan) maupun dinamis (misalnya, konteks waktu, cuaca saat itu, atau bahkan perubahan mood yang terdeteksi dari interaksi terkini) menjadi krusial (Ricci dkk., 2010 dan Falk, 2019).

Meskipun teknologi berkembang pesat, sektor pariwisata tetap dipengaruhi oleh faktor eksternal seperti kebijakan, tren ekonomi, sosial, dan lingkungan (Achmad dkk., 2023 dan Rumanti dkk., 2023). Banyak platform digital wisata masih bersifat statis (Choi dkk., 2021) dan kurang adaptif terhadap perubahan real-time seperti cuaca (Yoon & Choi, 2023), tren media sosial (Shuvo & Islam, 2024), atau kebijakan terbaru. Akibatnya, rekomendasi cenderung terfokus pada destinasi populer, sementara lokasi potensial yang kurang dikenal sering terabaikan (Barykin dkk., 2021).

Berdasarkan data DISPARBUDPORA Kabupaten Sumedang (2024), terlihat ketimpangan kunjungan wisata yang sangat signifikan. Dari 77 destinasi terdaftar, 10 destinasi utama seperti Menara Kujang Sapasang (306.333 kunjungan), OW Janspark (249.579 kunjungan), dan OW Cipanas Cileungsing (81.735 kunjungan) menyumbang 72,4% total kunjungan (1,31 juta wisatawan). Sementara itu, 67 destinasi lainnya hanya menarik 27,6% (501.768 wisatawan), dengan beberapa lokasi seperti OW Taman Empang Sindulang (2 kunjungan) dan OW Golempang (7 kunjungan) nyaris tak tersentuh. Destinasi budaya bernilai tinggi seperti Kampung Adat Cikondang (8.450 kunjungan) dan Museum Prabu Geusan Ulun (14.232 kunjungan) juga terabaikan, meskipun memiliki potensi sejarah yang kaya.

Ketimpangan ini diperparah oleh distribusi musiman yang timpang. Destinasi populer seperti Menara Kujang Sapasang ramai sepanjang tahun, sedangkan destinasi alam seperti Tahura Gunung Kunci (20.981 kunjungan) hanya mengalami peningkatan kunjungan terjadi pada Juli-Agustus dan Desember, didorong oleh musim liburan. Minimnya promosi digital turut memperburuk situasi: OW Pesona Taman Puspa dengan jumlah kunjungan 2.354 yang hanya memiliki ulasan di Google sebanyak 38 ulasan. Situasi ini mempertegas perlunya sistem rekomendasi adaptif yang tidak hanya mengandalkan popularitas, tetapi juga mengintegrasikan pendekatan keberagaman (*diversity*), yaitu strategi merekomendasikan destinasi dengan kategori, lokasi, dan tingkat popularitas beragam. Dengan demikian, sistem ini dapat memperluas cakupan destinasi, mengalihkan kunjungan dari destinasi *overcrowded*, sekaligus mendorong pemerataan ekonomi pariwisata melalui peningkatan eksposur destinasi potensial yang selama ini kurang terekspos.

Untuk menjembatani kesenjangan ini, sistem rekomendasi yang mampu beradaptasi dengan perubahan kondisi sangat dibutuhkan. Tidak hanya mengandalkan data historis, sistem ini harus mengintegrasikan informasi terbaru dari tren media sosial, ulasan terkini, hingga faktor lingkungan. Sebagai contoh, data tren media sosial bisa berupa kajian terhadap tagar yang sedang ramai seperti #ExploreSumedang, pengenalan konten viral seputar kegiatan atau makanan khas daerah, serta pengamatan pandangan masyarakat terhadap tempat wisata tertentu.

Ulasan terbaru dari pengguna di berbagai platform memberikan umpan balik langsung tentang standar fasilitas, tingkat kepadatan yang sebenarnya, atau perubahan pelayanan di suatu destinasi wisata. Di sisi lain, aspek lingkungan dapat mencakup informasi ramalan cuaca detail per jam atau peringatan awal mengenai kondisi alam dalam skala lokal.

Mengingat karakteristik data yang terus mengalir dan berubah dari beragam sumber ini, dibutuhkan penerapan metode pemrosesan aliran data (*streaming data processing*) (Alotaibi dkk., 2023). Pendekatan ini memungkinkan sistem melakukan analisis dan memberikan respons terhadap data secara langsung ketika informasi masuk, tanpa harus menunggu proses pengolahan terjadwal (*batch processing*) (Fragkoulis dkk., 2024 dan Alotaibi dkk., 2023). Hal ini menciptakan kemampuan penyesuaian rekomendasi yang lebih responsif dan sesuai dengan dinamika kondisi yang terjadi di lapangan (Fragkoulis dkk., 2024).

Dalam proses evaluasi, aspek akurasi prediksi, relevansi rekomendasi, serta keberagaman destinasi yang disarankan menjadi kunci utama. Menurut Akhadam dkk., (2025), pengukuran akurasi seperti *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE) diterapkan untuk mengevaluasi seberapa akurat sistem dalam memperkirakan penilaian atau ketertarikan pengguna pada sebuah destinasi, sementara mutu daftar rekomendasi dievaluasi menggunakan *Precision* dan *Recall* guna mengukur kapasitas sistem dalam menampilkan destinasi yang tepat sasaran, ditambah dengan *Normalized Discounted Cumulative Gain* (NDCG) yang berperan penting memastikan destinasi paling cocok tampil pada posisi paling atas. Sebagai pelengkap evaluasi dan penilaian pengalaman pengguna secara menyeluruh, dilakukan pengukuran terhadap *Diversity* guna menjamin keberagaman kategori wisata yang diusulkan (seperti alam, budaya, dan kuliner), *Novelty* untuk mengevaluasi kapabilitas sistem menampilkan destinasi baru atau 'harta karun tersembunyi' di luar yang sudah terkenal, serta *Coverage* untuk memastikan sistem mampu merekomendasikan seluruh rangkaian destinasi yang tersedia, tidak terbatas pada yang populer saja.

Beberapa penelitian di bidang rekomendasi pariwisata telah menunjukkan beragam kemajuan, namun aspek keberagaman (*diversity*) belum menjadi fokus utama. Misalnya, Puspasari (2022) menitikberatkan peningkatan akurasi dengan KNN dan SVM untuk rekomendasi pameran museum, sementara Chalkiadakis dkk., (2023) memadukan *Bayesian inference* dan *content-based similarity measures*, meraih *Precision* di atas 83% pada data pariwisata nyata. Adapun Yoon dan Choi, (2023) menerapkan pendekatan *context-aware* berbasis data *real-time* (cuaca, jarak, dsb.) serta beberapa model ML (KNN, SVM, dsb.) untuk rekomendasi destinasi, menghasilkan *Accuracy* 77,3%, micro-F1 0,773, dan macro-F1 0,415. Meski demikian, penelitian-penelitian tersebut umumnya tidak menyoroti keberagaman maupun cakupan (*coverage*), sehingga destinasi kurang populer cenderung belum diakomodasi secara optimal.

Penelitian-penelitian terdahulu telah membuka jalan menuju pengembangan sistem rekomendasi yang lebih cerdas. Namun, masih ada tantangan yang perlu diatasi, terutama dalam hal akurasi, keberagaman respons, dan integrasi data *real-time*. Sehingga masih perlu untuk melakukan penelitian untuk mengembangkan solusi alternatif dan inovasi dalam sistem rekomendasi adaptif yang mengoptimalkan keberagaman (*diversity*) dan cakupan (*coverage*) destinasi melalui integrasi data real-time seperti cuaca, tren media sosial, dan kondisi lingkungan. Dengan pendekatan ini, diharapkan terjadi redistribusi kunjungan wisatawan yang lebih adil, peningkatan eksposur destinasi potensial, serta peningkatan daya saing industri pariwisata melalui pengalaman wisatawan yang lebih optimal.

I.2 Rumusan Masalah

Berikut merupakan rumusan masalah yang akan dibahas dalam penelitian ini:

1. Bagaimana integrasi data real-time (cuaca, lalu lintas, penanggalan, dan tren media sosial) dapat meningkatkan keberagaman (*diversity*) dan cakupan (*coverage*) rekomendasi dibandingkan pendekatan statis, sekaligus tetap mempertahankan akurasi (*F1-score*) dan relevansi (*Precision, Recall*)?
2. Seberapa besar pengaruh modul streaming data real-time (cuaca, lalu lintas, penanggalan, dan tren media sosial) dalam mendorong pemerataan pariwisata, khususnya dengan mengangkat destinasi kurang terekspos, diukur melalui peningkatan *diversity*?
3. Sejauh mana sistem rekomendasi adaptif *real-time* dapat meningkatkan minat dan kunjungan wisatawan ke destinasi kurang populer, diukur melalui *Novelty* dengan metrik Popularity (Pop) dan hasil pengujian pengguna (*user testing*)?

I.3 Tujuan Penelitian

Tujuan utama penelitian ini adalah mengembangkan sistem rekomendasi adaptif real-time untuk pariwisata yang tidak hanya mempertahankan akurasi (*F1-score, Precision, Recall*) tetapi juga meningkatkan keragaman (*diversity*), cakupan (*coverage*), dan *novelty* destinasi, sehingga mendukung pemerataan pariwisata dengan mengangkat destinasi kurang terekspos dalam berbagai kondisi yang berubah-ubah (cuaca, kepadatan, tren media sosial). Beberapa tujuan spesifik yang dicapai dalam rangka mendukung tujuan utama adalah:

1. Mengembangkan modul integrasi data real-time (cuaca, lalu lintas, penanggalan, tren media sosial) dan memvalidasi dampaknya terhadap peningkatan keragaman (*diversity*) dan cakupan (*coverage*) rekomendasi, serta mempertahankan akurasi (*F1-score*) dan relevansi (*Precision, Recall*).
2. Menerapkan teknik streaming data *real-time* dalam sistem rekomendasi untuk menilai sejauh mana keragaman dan cakupan destinasi dapat ditingkatkan.

3. Merancang dan menguji sistem rekomendasi adaptif *real-time* yang diukur melalui metrik *novelty* serta *user testing*, untuk menilai dampaknya terhadap ketertarikan wisatawan terhadap destinasi non-populer.

I.4 Ruang Lingkup Masalah

Untuk meningkatkan fokus dari penelitian, diperlukan beberapa batasan dalam penyelesaian masalah penelitian. Berikut merupakan batasan masalah dari penelitian ini.

1. Penggunaan data *real-time* meliputi pembaruan cuaca, kondisi lalu lintas, penanggalan, dan tren media sosial.
2. Penelitian difokuskan pada destinasi wisata di Kabupaten Sumedang.
3. Analisis dilakukan terhadap tingkat keberagaman dan cakupan rekomendasi destinasi.
4. Evaluasi performa sistem mencakup metrik *F1-score*, *Precision*, *Recall*, *diversity* dan *coverage*.
5. Evaluasi berbasis pengguna dilakukan melalui *user testing* dengan pengamatan langsung terhadap interaksi pengguna.

I.5 Kebaruan dan Kontribusi

Adapun kebaruan dan kontribusi penelitian ini adalah sebagai berikut.

1. Penelitian ini mengusulkan sistem rekomendasi yang menggabungkan beragam sumber data *real-time* (cuaca, kondisi lalu lintas, penanggalan, dan tren media sosial) melalui teknologi *data streaming* dan *machine learning*. Kebaruan terletak pada kemampuan sistem untuk secara dinamis memperbarui rekomendasi berdasarkan perubahan kondisi lingkungan, mengatasi keterbatasan sistem statis.
2. Berbeda dengan kebanyakan studi terdahulu yang berfokus pada akurasi semata, penelitian ini menambahkan modul *bias correction* pada tahap ranking. Strategi ini bertujuan meningkatkan *diversity* dan *coverage*, sehingga destinasi kurang populer namun relevan dapat diikutsertakan dalam rekomendasi.

3. Sistem ini diharapkan dapat meningkatkan minat dan kunjungan wisatawan ke destinasi kurang populer. Evaluasi melalui metrik *novelty* dan *user testing* memungkinkan pengukuran seberapa efektif sistem dalam mempromosikan destinasi baru. Penelitian ini memberikan kontribusi dalam mengembangkan teknologi rekomendasi wisata yang adaptif, beragam, dan mendukung pemerataan ekonomi pariwisata.

I.6 Sistematika Penulisan

Dalam penyusunan tesis ini, penulis terbagi menjadi lima bab dengan penjabaran untuk masing-masing bab sebagai berikut.

BAB I Pendahuluan

Bab ini membahas mengenai latar belakang masalah, rumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, dan sistematika penulisan yang digunakan.

BAB II Tinjauan Pustaka

Bab ini membahas mengenai konsep dasar serta teori-teori yang berkaitan dengan topik penelitian dan hal-hal yang berguna dalam proses analisis permasalahan.

BAB III Metodologi Penelitian

Bab ini berisi tentang metodologi penelitian yang digunakan dalam melakukan penelitian serta dijelaskan metode yang dilakukan setiap tahapan mulai dari awal sampai dengan akhir.

BAB IV Analisis Dan Desain Awal

Bab ini berisi analisis kebutuhan fungsional dan non-fungsional pada sistem yang akan dibuat. Pada bagian ini juga menjelaskan rancangan awal yang akan dibuat pada penelitian ini.

BAB II. Tinjauan Pustaka

Tinjauan pustaka dalam penelitian ini membahas teori-teori yang mendasari pengembangan sistem rekomendasi pariwisata adaptif. Bagian ini mencakup kajian teori, penelitian terdahulu, serta kerangka pemikiran yang menjadi dasar dalam perancangan sistem.

II.1 Sistem rekomendasi

Sistem rekomendasi (*Recommender Systems* - RSs) adalah teknologi perangkat lunak yang dibuat untuk menyarankan item yang sesuai bagi pengguna (Ricci dkk., 2010). Sistem ini membantu dalam berbagai proses pengambilan keputusan, seperti menentukan produk yang akan dibeli, memilih musik yang sesuai, atau menemukan berita daring yang menarik untuk dibaca (Alfaifi, 2024). Fungsi utama sistem rekomendasi adalah menyeleksi dan mengolah data dalam jumlah besar untuk menyuguhkan informasi yang relevan dan disesuaikan dengan preferensi masing-masing pengguna (Ricci dkk., 2010). Hal ini berperan penting dalam mereduksi dampak kelebihan informasi (Kamal dkk., 2024).

Agar dapat beroperasi dengan maksimal, sistem rekomendasi membutuhkan data yang komprehensif dan melingkupi beragam aspek. Kategori data primer mencakup: informasi pengguna, baik yang disampaikan secara eksplisit seperti rating dan komentar, maupun yang diperoleh secara implisit melalui perilaku dan pola penggunaan (Ricci dkk., 2010), data demografis seperti umur, gender, dan negara asal untuk memperkaya profil pengguna (Ricci dkk., 2010 & Falk, 2019), informasi tentang item termasuk karakteristik dan penjelasan produk, data kontekstual yang menggambarkan situasi ketika interaksi berlangsung seperti tipe perangkat, waktu akses, posisi geografis, dan status pengguna apakah sendirian atau bersama kelompok, serta rekaman interaksi sebelumnya dan tanggapan yang menjadi fondasi bagi berbagai pendekatan algoritmik, terutama dalam metode Collaborative Filtering.

II.2 Aplikasi dan implementasi

Sistem rekomendasi diterapkan dalam produk berwujud dan tidak berwujud serta industri perawatan layanan (Alfaifi, 2024), hiburan, rekomendasi situs web *e-learning*, rekomendasi surat kabar, dan filter email (Falk, 2019), aplikasi *e-commerce*, sistem peer-to-peer (Kamal dkk., 2024) dan *Blockchain* (Sachi Nandan Mohanty, 2020).

Dalam sektor pariwisata, sistem rekomendasi telah berkembang pesat dengan implementasi yang sangat beragam. Menurut Solano-Barliza dkk., (2024), penerapan sistem rekomendasi di industri pariwisata mencakup usulan tujuan wisata, aplikasi ponsel pintar sebagai asisten perjalanan, sistem terintegrasi berbasis data lokasi, paket wisata yang dipersonalisasi, layanan e-tourism terpadu, situs-situs bersejarah dan budaya, serta infrastruktur pengisian kendaraan listrik yang berperan dalam mewujudkan konsep pariwisata berkelanjutan dan teknologi cerdas.

Walaupun mengalami perkembangan yang pesat, sistem rekomendasi pariwisata tetap berhadapan dengan berbagai kendala, seperti:

1. Keterbatasan adaptasi real-time merupakan masalah pokok dimana sistem tidak dapat merespons dengan baik terhadap dinamika perubahan seperti kondisi cuaca, popularitas media sosial, atau regulasi baru, meski beberapa sistem sudah mempertimbangkan faktor waktu namun masih rentan dalam kondisi yang cepat berubah (Noorian, 2024).
2. Preferensi pada destinasi populer (*Long Tail Problem*) terlihat dari kecenderungan sistem khususnya *Collaborative Filtering* memprioritaskan tempat wisata populer dengan rating tinggi, mengakibatkan lokasi potensial yang belum terkenal menjadi terabaikan (Ricci dkk., 2010).
3. Kendala *Cold Start* dan data terbatas (*Sparsity*) terjadi saat sistem sulit memberikan saran untuk destinasi atau user baru akibat minimnya rekam jejak, yang semakin memperburuk pengabaian terhadap destinasi kurang populer (Suhaim & Berri, 2021).
4. Personalisasi yang terbatas pada level individu dan kurangnya data demografis yang menghambat kemampuan kustomisasi karena sejumlah model tidak memasukkan karakteristik demografis pengguna (Kumar dkk., 2024).

II.3 Klasifikasi Sistem Rekomendasi

Dalam pengembangannya, sistem rekomendasi terklasifikasi menjadi beberapa teknik. Beberapa teknik yang dapat digunakan dalam sistem rekomendasi yang ditemukan pada penelitian Solano-Barliza dkk., (2024) yaitu *Collaborative filtering*, *Content-based filtering*, *Location-based*, *Context-aware*, *Hybrid*.

Klasifikasi ini bukan sekadar mengenali berbagai macam pendekatan, tetapi juga mengungkapkan keunggulan dan kelemahan yang melekat pada setiap metode, serta cara beberapa di antaranya saling mendukung untuk menciptakan sistem yang lebih kuat dan fleksibel.

II.3.1 *Collaborative filtering*

Model ini bekerja dengan menganalisis pola penilaian pengguna terhadap item yang sama untuk menentukan kesamaan preferensi mereka (Falk, 2019). Pendekatan ini memungkinkan sistem untuk memberikan rekomendasi yang dipersonalisasi dengan secara optimal memanfaatkan data yang tersimpan guna mengidentifikasi keterkaitan antara berbagai domain. Dari analisis ini, sistem menghasilkan daftar item yang paling sesuai dengan selera pengguna berdasarkan evaluasi dari kelompok pengguna dengan preferensi serupa.

1. *Memory-based*

Pada pendekatan ini, sistem menganalisis matriks utilitas untuk memahami pola penggunaan, lalu memberikan rekomendasi dengan mencocokkan preferensi pengguna dengan data yang ada dalam matriks tersebut (Sachi Nandan Mohanty, 2020).

2. *Model-based*

Model ini merepresentasikan hubungan antara pengguna dan item melalui matriks utilitas, yang dipecah menjadi dua matriks, yaitu A untuk pengguna dan B untuk item (Jie Tang dkk., 2013). Teknik seperti *Singular Value Decomposition* (SVD) dan *Principal Component Analysis* (PCA) digunakan dalam proses dekomposisi matriks. Sistem kemudian menghitung rating untuk setiap item dan merekomendasikan produk dengan nilai tertinggi (Sachi Nandan Mohanty, 2020). Pendekatan ini sangat efektif ketika jumlah data yang tersedia sangat besar.

II.3.2 *Content-based filtering*

Content-based filtering merupakan pendekatan dalam sistem rekomendasi yang menyarankan item berdasarkan karakteristik atau informasi dari kontennya sendiri (Sachi Nandan Mohanty, 2020). Namun, metode ini memiliki kekurangan, yakni tidak mampu merekomendasikan item yang tidak memiliki keterkaitan langsung dengan konten yang telah tersedia (Sachi Nandan Mohanty, 2020). Model ini umumnya diterapkan dalam layanan yang merekomendasikan teks atau artikel, tetapi kurang optimal dalam menangani item baru yang belum memiliki informasi relevan dalam sistem (Ricci dkk., 2010).

II.3.3 *Location-based*

Sistem rekomendasi berbasis lokasi memanfaatkan data geografis untuk menyajikan rekomendasi tempat atau aktivitas yang sesuai dengan minat pengguna (Shafqat dan Byun, 2020). Menurut Solano-Barliza dkk., (2024) terdapat dua kategori utama dalam sistem ini: sistem rekomendasi tempat mandiri dan sistem rekomendasi tempat berurutan. Sistem rekomendasi tempat mandiri menyesuaikan rekomendasi dengan preferensi pengguna, seperti restoran, hotel, atau destinasi wisata. Sementara itu, sistem rekomendasi tempat berurutan menawarkan informasi mengenai rute perjalanan yang populer, perkiraan waktu yang dibutuhkan, serta estimasi biaya perjalanan.

II.3.4 *Context-aware*

Sistem rekomendasi berbasis konteks dirancang untuk meningkatkan relevansi dan personalisasi rekomendasi dengan memperhitungkan faktor kontekstual dalam aktivitas pengguna (Shafqat dan Byun, 2020). Sistem ini terbagi ke dalam beberapa model, termasuk *collaborative filtering*, *content-based filtering*, pendekatan grafis, serta model hibrida (Solano-Barliza dkk., 2024). Pada penelitian Suhaim dan Berri (2021) terdapat informasi kontekstual dalam sistem ini diklasifikasikan berdasarkan faktor individu, jenis aktivitas, aspek temporal, lokasi geografis, dan komunikasi. Elemen yang diperhitungkan mencakup lokasi pengguna, waktu penggunaan, serta status koneksi jaringan.

II.3.5 Hybrid

Sistem rekomendasi hibrida merupakan pendekatan yang mengombinasikan beberapa metode untuk menghasilkan rekomendasi yang lebih akurat (Fararni dkk., 2021). Dengan mengintegrasikan berbagai teknik, sistem ini mampu mengatasi kelemahan dari masing-masing metode secara lebih efektif (Sachi Nandan Mohanty, 2020).

II.4 Metrik Evaluasi

Evaluasi pada suatu sistem rekomendasi diperlukan untuk menentukan seberapa efektif kinerja dari sistem rekomendasi tersebut. Menurut Ricci dkk. (2010) keluaran dari sistem rekomendasi memiliki empat kemungkinan hasil untuk item yang direkomendasikan dan tersembunyi, seperti yang ditunjukkan pada tabel II.1

Table II.1 Klasifikasi kemungkinan hasil rekomendasi

	Direkomendasi	Tidak Direkomendasikan
Digunakan	True-Positive (tp)	False-Negative (fn)
Tidak Digunakan	False-Positive (fp)	True-Negative (tn)

II.4.1 Metrik Akurasi Prediksi

1. MAE

Mengukur rata-rata nilai absolut dari perbedaan antara peringkat yang diprediksi oleh sistem dan peringkat sebenarnya dari pengguna (Choi dkk., 2021). Nilai MAE yang lebih rendah menunjukkan kinerja yang lebih baik, rentang nilai MAE adalah $[0, +\infty]$ (Sachi Nandan Mohanty, 2020). Perhitungan nilai *Mean Estimation Accuracy* (MEA) dilakukan berdasarkan Persamaan (II.1) berikut:

$$MAE = \frac{1}{N} \sum_{i=1}^N |r_{ui} - \tilde{r}_{ui}| \quad (II.1)$$

dengan:

r_{ui} = Rating aktual dari pengguna u untuk item i .

\tilde{r}_{ui} = Rating yang diprediksi dari pengguna u untuk item i .

N = Jumlah item dalam daftar rekomendasi (panjang daftar yang dievaluasi, misal $N=10$ untuk Top-10 rekomendasi).

2. RMEA

Mengukur akar kuadrat dari rata-rata kuadrat perbedaan antara peringkat yang diprediksi dan peringkat sebenarnya (Falk, 2019). RMSE memberikan bobot lebih besar pada kesalahan prediksi yang besar. Nilai RMSE yang lebih rendah juga menunjukkan kinerja yang lebih baik, rentang nilai RMSE adalah $[0, +\infty]$ (Sachi Nandan Mohanty, 2020). Perhitungan nilai RMSE dilakukan berdasarkan Persamaan (II.2) berikut:

$$RMAE = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_{ui} - \tilde{r}_{ui})^2} \quad (II.2)$$

dengan:

r_{ui} = Rating aktual dari pengguna u untuk item i .

\tilde{r}_{ui} = Rating yang diprediksi dari pengguna u untuk item i .

N = Jumlah item dalam daftar rekomendasi (panjang daftar yang dievaluasi, misal $N=10$ untuk Top-10 rekomendasi).

II.4.2 Metrik Akurasi Klasifikasi

1. Precision

Proporsi item yang direkomendasikan yang relevan bagi pengguna (Falk, 2019). Dihitung sebagai jumlah item yang direkomendasikan dengan benar dibagi dengan jumlah total item yang direkomendasikan (Sachi Nandan Mohanty, 2020). Presisi yang tinggi menunjukkan bahwa sebagian besar item yang direkomendasikan benar-benar relevan, sehingga nilai yang lebih tinggi adalah lebih baik. Rentang nilai Precision adalah $[0,1]$. Perhitungan nilai *Precision* dilakukan berdasarkan Persamaan (II.3) berikut:

$$Precision = \frac{|C_u \cap R_u|}{|R_u|} \quad (II.3)$$

dengan:

C_u = Daftar item yang dikonsumsi oleh pengguna.

R_u = Daftar rekomendasi untuk pengguna u .

2. *Recall*

Proporsi item yang relevan yang berhasil direkomendasikan oleh sistem. Dihitung sebagai jumlah item yang direkomendasikan dengan benar dibagi dengan jumlah total item yang relevan (Ricci dkk., 2010). Recall yang tinggi menunjukkan bahwa sistem berhasil menemukan sebagian besar item yang relevan, sehingga nilai yang lebih tinggi adalah lebih baik. Rentang nilai Recall adalah $[0,1]$ (Sachi Nandan Mohanty, 2020). Perhitungan nilai *Recall* dilakukan berdasarkan Persamaan (II.4) berikut:

$$Recall = \frac{|C_u \cap R_u|}{|C_u|} \quad (II.4)$$

dengan:

C_u = Daftar item yang dikonsumsi oleh pengguna.

R_u = Daftar rekomendasi untuk pengguna u .

3. *F1-score*

Rata-rata harmonik tertimbang dari presisi dan recall (Ricci dkk., 2010). *F1-score* memberikan ukuran tunggal yang menyeimbangkan *precision* dan *recall*, sehingga nilai yang lebih tinggi adalah lebih baik. Rentang nilai F1-score adalah $[0,1]$ (Sachi Nandan Mohanty, 2020). Perhitungan nilai *F1-score* dilakukan berdasarkan Persamaan (II.5) berikut:

$$F1\ Score = \frac{2PR}{P + R} \quad (II.5)$$

dengan:

P = Precision

R = Recall

II.4.3 Metrik Akurasi Ranging

1. *Pearson correlation*

Korelasi Pearson digunakan untuk menilai seberapa kuat hubungan linear antara penilaian yang diprediksi dan penilaian aktual milik pengguna. Nilai yang lebih tinggi (mendekati 1 untuk korelasi positif, atau mendekati -1 untuk korelasi negatif, tergantung pada konteks yang diinginkan, namun umumnya target "Higher" untuk korelasi positif) menunjukkan kinerja yang lebih baik.

Rentang nilai korelasi Pearson adalah $[-1,1]$ (Akhadam dkk., 2025). Perhitungan nilai *Pearson correlation* dilakukan berdasarkan Persamaan (II.6) berikut:

$$PC = \frac{\sum_{i=1}^N (r_{ui} - \bar{r}_u) (\tilde{r}_{ui} - \tilde{\bar{r}}_u)}{\sum_{i=1}^N (r_{ui} - \bar{r}_u)^2 (\tilde{r}_{ui} - \tilde{\bar{r}}_u)^2} \quad (II.6)$$

dengan:

r_{ui} = Rating aktual dari pengguna u untuk item i .

\tilde{r}_{ui} = Rating yang diprediksi dari pengguna u untuk item i .

\bar{r}_u = Rata-rata rating aktual dari pengguna u .

$\tilde{\bar{r}}_u$ = Rata-rata rating yang diprediksi dari pengguna u .

N = Jumlah item dalam daftar rekomendasi (panjang daftar yang dievaluasi, misal $N=10$ untuk Top-10 rekomendasi).

2. *Spearman correlation*

Korelasi Spearman menitikberatkan pada urutan peringkat (ranking) penilaian, mengukur hubungan linear antar peringkat penilaian yang diprediksi dan yang sebenarnya. Nilai yang lebih tinggi (mendekati 1 untuk korelasi positif) menunjukkan kinerja yang lebih baik. Rentang nilai korelasi Spearman adalah $[-1,1]$ (Akhadam dkk., 2025). Perhitungan nilai *Spearman correlation* dilakukan berdasarkan Persamaan (II.7) berikut:

$$SC = \frac{\sum_{i=1}^N (r_i - \bar{r}_u) (\tilde{r}_i - \tilde{\bar{r}}_u)}{\sum_{i=1}^N (r_i - \bar{r}_u)^2 (\tilde{r}_i - \tilde{\bar{r}}_u)^2} \quad (II.7)$$

dengan:

r_i = Peringkat item i dalam daftar rekomendasi.

\bar{r}_u = Rata-rata rating yang diprediksi dari pengguna u .

$\tilde{\bar{r}}_u$ = Rata-rata rating yang diprediksi dari pengguna u .

N = Jumlah item dalam daftar rekomendasi (panjang daftar yang dievaluasi, misal $N=10$ untuk Top-10 rekomendasi).

II.4.4 Metrik Berbasis Peringkat

1. *Normalized Discounted Cumulative Gain* (NDCG)

Menilai kegunaan item dalam daftar rekomendasi berdasarkan posisinya, di mana item relevan di posisi atas lebih bernilai. Nilai NDCG yang lebih tinggi menunjukkan kinerja peringkat yang lebih baik. Rentang nilai NDCG adalah $[0,1]$ (Akhadam dkk., 2025). Perhitungan nilai NDCG dilakukan berdasarkan Persamaan (II.8) berikut:

$$NDGC = \frac{\sum_{i=1}^b R_i + \sum_{i=b+1}^N \frac{R_i}{\log_b r_i}}{DCG_{max}} \quad (II.8)$$

dengan:

R_i = Relevansi item i dalam daftar rekomendasi.

b = Basis dari logaritma.

r_i = Peringkat item i dalam daftar rekomendasi.

DCG_{max} = Gain maksimum yang mungkin didapatkan.

N = Jumlah item dalam daftar rekomendasi (panjang daftar yang dievaluasi, misal $N=10$ untuk Top-10 rekomendasi).

2. *Recovery Rate*

Recovery Rate menilai seberapa akurat sistem dalam menyusun item pada urutan yang tepat. Menurut Akhadam dkk. (2025), nilai Recovery Rate yang lebih rendah menunjukkan kinerja yang lebih baik. Rentang nilai RR adalah $[0,1]$. Perhitungan nilai *Recovery Rate* dilakukan berdasarkan Persamaan (II.9) berikut:

$$RR = \frac{1}{|N_r|} \sum_{i \in N_r} \frac{r_i}{C_i} \quad (II.9)$$

dengan:

C_i = Jumlah item kandidat yang direkomendasikan kepada pengguna u .

N_r = Kumpulan item relevan dalam daftar rating aktual pengguna u .

r_i = Peringkat item i dalam daftar rekomendasi.

3. *Rank Biased Precision* (RBP)

Rank Biased Precision (RBP), mirip dengan DCG, mengevaluasi daftar rekomendasi dengan memberikan bobot lebih pada item relevan yang memiliki peringkat lebih tinggi. Perbedaannya terletak pada fungsi diskon yang digunakan: RBP menerapkan fungsi diskon geometris, sedangkan DCG menggunakan fungsi logaritmik (Akhadam dkk., 2025). Perhitungan nilai RBP dilakukan berdasarkan Persamaan (II.10) berikut:

$$RBP = (1 - p) \sum_{i=1}^N (R_i \times p^{i-1}) \quad (II.10)$$

dengan:

R_i = Relevansi item i dalam daftar rekomendasi.

N = Jumlah item dalam daftar rekomendasi (panjang daftar yang dievaluasi, misal $N=10$ untuk Top-10 rekomendasi).

II.4.5 Metrik Lainnya

1. *Diversity*

Keberagaman mengukur seberapa bervariasi item-item yang disarankan dalam daftar rekomendasi. Terdapat dua dimensi, yaitu *intra-diversity* yang menilai perbedaan antar item yang ditawarkan kepada seorang pengguna, dan *inter-diversity* yang mengukur sejauh mana daftar rekomendasi berbeda di antara semua pengguna (Akhadam dkk., 2025). Perhitungan nilai *intra-diversity* dilakukan berdasarkan Persamaan (II.11) berikut:

$$ID = \frac{1}{N(N-1)} \times \frac{\sum_{i \neq j} s(I_i, I_j)}{m} \quad (II.11)$$

dengan:

$s(I_i, I_j)$ = Kemiripan antara item i dan j .

m = Jumlah total pengguna dalam sistem.

N = Jumlah item dalam daftar rekomendasi (panjang daftar yang dievaluasi, misal $N=10$ untuk Top-10 rekomendasi).

Untuk metrik *Intra-diversity* ini, agar nilai yang lebih tinggi mencerminkan keberagaman yang lebih baik, maka $s(I_i, I_j)$ dalam Persamaan (II.11) dipahami sebagai representasi ukuran ketidakmiripan antar item. Ini dapat dicapai, misalnya, dengan menghitung $s(I_i, I_j)$ sebagai $1 - \text{skor kemiripan } (I_i, I_j)$. Dengan demikian, nilai ID yang lebih tinggi akan menunjukkan bahwa item-item dalam daftar rekomendasi pengguna lebih beragam.

Nilai *Inter-diversity* (H) yang lebih tinggi adalah lebih baik. Ini menunjukkan bahwa daftar rekomendasi yang diterima oleh pengguna yang berbeda semakin bervariasi, artinya sistem tidak hanya menyarankan item yang itu-itu saja kepada semua orang. Perhitungan nilai *inter-diversity* dilakukan berdasarkan Persamaan (II.12) berikut:

$$H = \frac{1}{m(m-1)} \sum_{u=1}^m \sum_{u' \neq u} 1 - \left(\frac{Q_{uu'}}{N}\right) \quad (\text{II.12})$$

dengan:

$Q_{uu'}$ = Jumlah item yang sama dalam daftar rekomendasi pengguna u dan u' .

m = Jumlah total pengguna dalam sistem.

N = Jumlah item dalam daftar rekomendasi (panjang daftar yang dievaluasi, misal $N=10$ untuk Top-10 rekomendasi).

2. Coverage

Sebuah sistem rekomendasi sebaiknya mampu menjangkau berbagai macam item. Oleh karena itu, konsep *Coverage* diperkenalkan, yang menyatakan proporsi item dari katalog yang bisa direkomendasikan. Nilai *Catalogue Coverage* (C) yang lebih tinggi menunjukkan kinerja yang lebih baik. Rentang nilainya adalah dari $[0,1]$ (Akhadam dkk., 2025). Perhitungan nilai *Catalogue Coverage* dilakukan berdasarkan Persamaan (II.13) berikut:

$$C = \frac{I_r}{n} \quad (\text{II.13})$$

dengan:

I_r = Jumlah total item yang berbeda dalam daftar top-N pengguna.

n = Jumlah total item dalam sistem.

Menguji seberapa sering sebuah item disarankan kepada pengguna. Nilai *Entropy Coverage* (EC) yang lebih tinggi menunjukkan distribusi rekomendasi item yang lebih merata dan beragam antar pengguna. Rentang nilainya adalah $[0, +\infty]$. Perhitungan nilai *Entropy Coverage* dilakukan berdasarkan Persamaan (II.14) berikut:

$$EC = - \sum_{i=1}^m p_i \log_2 p_i \quad (II.14)$$

dengan:

p_i = Persentase daftar rekomendasi yang mengandung item i .

m = Jumlah total pengguna dalam sistem.

3. *Novelty*

Novelty berfokus pada seberapa banyak item baru yang muncul dalam daftar rekomendasi. Untuk mengukur *Novelty*, metrik *Popularity* (Pop) digunakan. Nilai *Popularity* yang lebih rendah menunjukkan *Novelty* yang lebih tinggi, karena sistem merekomendasikan item-item yang kurang populer. Rentang nilai *Popularity* adalah $[0,1]$ (Akhadam dkk., 2025). Perhitungan nilai *Novelty* dilakukan berdasarkan Persamaan (II.15) berikut:

$$Pop = \frac{1}{m \times N} \sum_{u=1}^m \sum_{i=1}^N d_i \quad (II.15)$$

dengan:

m = Jumlah total pengguna dalam sistem.

N = Jumlah item dalam daftar rekomendasi (panjang daftar yang dievaluasi, misal $N=10$ untuk Top-10 rekomendasi).

d_i = Derajat item i .

di mana $Pop(i)$ adalah skor popularitas dari item i (jumlah total interaksi yang diterima item i dari seluruh pengguna dalam data historis).

II.5 *Machine Learning*

Machine learning merupakan cabang dari kecerdasan buatan yang fokus pada pembuatan algoritma guna melatih mesin agar mampu menjalankan tugas spesifik (Yazici dkk., 2023). Proses ini dimulai dengan pengumpulan dataset, kemudian menganalisis hubungan antar data, mengidentifikasi pola-pola, dan menerapkan algoritma sehingga mesin dapat mengubah data input menjadi output yang diinginkan berdasarkan pola yang terdeteksi (Singh dkk., 2023).

Peran *Machine Learning* dalam sistem rekomendasi adalah mengidentifikasi pola, memprediksi preferensi pengguna, dan menyajikan item yang paling sesuai (Sachi Nandan Mohanty, 2020). Penerapan ML memungkinkan sistem rekomendasi untuk memahami perilaku pengguna secara dinamis dan menghasilkan saran yang lebih personal dan relevan (Yoon & Choi, 2023). Rekomendasi yang dihasilkan oleh sistem *Machine Learning* ini bertujuan untuk menciptakan rekomendasi yang bermakna bagi sejumlah pengguna terkait item atau produk yang mungkin menarik minat mereka (Sachi Nandan Mohanty, 2020).

Penerapan metode *Machine Learning* dalam sistem rekomendasi dapat dikategorikan berdasarkan proses pembelajarannya menjadi tiga kelompok utama: *supervised learning*, *unsupervised learning*, dan *reinforcement learning* (Yazici dkk., 2023).

II.5.1 *Supervised Learning*

Dalam jenis pembelajaran ini, data yang digunakan untuk melatih model dilabeli dengan kelas dan hasil yang diharapkan (Pugliese dkk., 2021). Beberapa contoh algoritma *supervised learning* meliputi:

1. *Naïve Bayes classification*

Algoritma ini didasarkan pada teorema Bayes dan asumsi "naif" tentang independensi bersyarat antara semua pasangan variabel. Dalam klasifikasi, *Naïve Bayes* menghitung probabilitas posterior untuk setiap kelas, dengan mempertimbangkan sekumpulan variabel independen (fitur) (Akhadam dkk., 2025). Kelas dengan probabilitas tertinggi kemudian dipilih sebagai prediksi (Sachi Nandan Mohanty, 2020).

2. *Linear dan Logistic Regression*

Regresi linier biasanya digunakan ketika variabel dependen bersifat numerik (kontinu), sedangkan regresi logistik digunakan ketika variabel dependen bersifat kategorikal (biasanya biner). Regresi linier mencoba untuk memodelkan hubungan linier antara variabel input dan output (Sarker, 2021). Regresi logistik mengukur probabilitas kelas tertentu menggunakan fungsi logistik (sigmoid), yang memetakan input ke nilai antara 0 dan 1. Peningkatan variabel independen akan meningkatkan odds output pada tingkat yang konstan (Falk, 2019).

3. *Support Vector Machines (SVMs)*

Algoritma supervised yang dapat digunakan untuk klasifikasi dan regresi. Namun, ini sering digunakan sebagai pengklasifikasi diskriminatif. SVM bekerja dengan menemukan hyperplane optimal yang memisahkan data ke dalam kelas-kelas yang berbeda dalam ruang berdimensi tinggi (Sachi Nandan Mohanty, 2020). Margin terbesar antara hyperplane dan titik data terdekat (support vector) menghasilkan generalisasi yang lebih baik (Akhadam dkk., 2025). Kernel (seperti RBF kernel) dapat digunakan untuk membuat batas keputusan non-linier jika kelas tidak dapat dipisahkan secara linier (Sarker, 2021).

4. *Random Forests*

Random Forest adalah algoritma *supervised classification* yang merupakan metode ensemble learning. Ini membangun "hutan" yang terdiri dari banyak pohon keputusan. Semakin banyak pohon dalam hutan, semakin robust hutan tersebut. Untuk klasifikasi, *Random Forest* menggabungkan prediksi dari setiap pohon dan memilih kelas dengan suara terbanyak. Untuk regresi, rata-rata prediksi dari setiap pohon diambil. Algoritma ini menggunakan teknik *bootstrap aggregating* (bagging) dan pemilihan fitur acak saat membangun setiap pohon (Sachi Nandan Mohanty, 2020).

5. *Neural Networks*

Jaringan saraf tiruan (*Neural Networks*) adalah subset dari *machine learning* yang terinspirasi oleh struktur dan fungsi otak manusia (Akhadam dkk., 2025). Mereka terdiri dari unit pemrosesan yang saling berhubungan yang disebut neuron tiruan (Janiesch dkk., 2021).

6. *Decision Trees*

Pohon keputusan (*Decision Trees*) adalah model klasifikasi atau regresi yang direpresentasikan dalam struktur seperti pohon (Ricci dkk., 2010). Setiap node internal dalam pohon mewakili pengujian pada suatu atribut, setiap cabang mewakili hasil dari pengujian tersebut, dan setiap leaf node mewakili keputusan atau prediksi akhir (Sachi Nandan Mohanty, 2020).

7. *K-Nearest Neighbor (KNN)*

KNN adalah algoritma *supervised* yang digunakan untuk klasifikasi dan regresi (Sachi Nandan Mohanty, 2020). Ini adalah algoritma "lazy learning" atau "instance-based learning" karena tidak membangun model eksplisit selama fase pelatihan, tetapi menyimpan semua data pelatihan. Untuk mengklasifikasikan instance baru, KNN mencari k instance terdekat (tetangga) dalam data pelatihan berdasarkan ukuran jarak (misalnya, *Euclidean distance*, *Cosine similarity*, *Pearson correlation*) (Sarker, 2021). Kelas mayoritas di antara k tetangga terdekat kemudian ditetapkan sebagai kelas untuk instance baru (Sachi Nandan Mohanty, 2020).

II.5.2 *Unsupervised Learning*

Dalam jenis pembelajaran ini, data yang digunakan tidak dilabeli (*data are not labeled*) (Pugliese dkk., 2021). Model belajar untuk menemukan pola, struktur tersembunyi, atau pengelompokan dalam data tanpa panduan dari label yang telah ditentukan (Janiesch dkk., 2021). Beberapa contoh algoritma unsupervised learning meliputi:

1. *K-means clustering*

Algoritma pembelajaran tanpa pengawasan (*unsupervised machine learning algorithm*) (Falk, 2019). Ini merupakan metode partisional (*partitioning method*) yang bertujuan untuk membagi dataset yang terdiri dari N item ke dalam k subset (*cluster*) yang saling terpisah. Setiap *cluster* didefinisikan oleh anggota-anggotanya dan sebuah centroid (Ricci dkk., 2010).

2. *Principal Component Analysis (PCA)*

Merupakan metode reduksi dimensi yang memproyeksikan data dengan dimensi tinggi ke ruang berdimensi lebih rendah, sekaligus berupaya mempertahankan informasi paling esensial (Pugliese dkk., 2021 dan Sachi Nandan Mohanty, 2020).

3. *Self-Organizing Map (SOM)*

Teknik *unsupervised learning* yang menggunakan jaringan saraf untuk merepresentasikan data berdimensi tinggi ke dalam peta 2D, mencapai reduksi dimensi (Sachi Nandan Mohanty, 2020 dan Sarker, 2021).

II.6 Basis Data

Basis data merupakan kumpulan data yang disimpan secara terstruktur dan persisten, sehingga dapat diakses serta dikelola. Pengelolaannya umumnya dilakukan oleh perangkat lunak bernama *Database Management System (DBMS)*. Berbagai jenis basis data tersedia untuk memenuhi beragam kebutuhan, misalnya basis data relasional yang menggunakan tabel, serta basis data *Not Only SQL (NoSQL)* dengan struktur yang lebih fleksibel (Taipalus, 2024). Kinerja basis data menjadi hal krusial, sehingga berbagai teknik optimasi diterapkan guna meningkatkan throughput dan menurunkan latensi (Huang dkk., 2023).

Dalam sistem rekomendasi, basis data berperan penting sebagai landasan infrastruktur data yang mendukung proses pengumpulan, penyimpanan, pemrosesan, dan analisis informasi penting untuk menghasilkan rekomendasi yang relevan dan terpersonalisasi bagi pengguna (Alfaifi, 2024). Tanpa basis data yang terstruktur dan terisi dengan baik, sistem rekomendasi tidak akan mampu berfungsi dan menghasilkan saran yang bermakna (Falk, 2019).

II.6.1 Jenis Basis Data

1. Basis Data Relasional

Sistem ini menggunakan model relasional untuk menyimpan data dalam tabel-tabel yang saling terkait, serta memanfaatkan *Structured Query Language* (SQL) untuk berinteraksi dengan data. Fokus utamanya adalah konsistensi dan integritas data. Desain basis data mengutamakan normalisasi guna mengurangi redundansi dan mencegah anomali. Beberapa *Relational Database Management System* (RDBMS) populer di antaranya PostgreSQL, Oracle Database, MySQL, SQL Server, Access, dan DB/2. Dalam basis data relasional, pemisahan antara data dan DBMS sulit dilakukan tanpa mengeksport data ke format lain (Taipalus, 2024).

2. Basis Data NoSQL

NoSQL adalah istilah umum yang mencakup berbagai model data (key-value, dokumen, wide-column, graph) tanpa struktur kaku sebagaimana pada RDBMS. Tiap DBMS NoSQL biasanya memiliki bahasa kueri khusus, yang umumnya kurang ekspresif dibandingkan SQL. Pendekatan konsistensi yang diterapkan pun berbeda: NoSQL lebih mengutamakan kinerja dan ketersediaan, sering menyimpan data secara berlebihan guna memudahkan kueri dan meningkatkan kecepatan, meskipun dapat menimbulkan inkonsistensi. Fokus utamanya adalah performa dan skalabilitas melalui replikasi serta sharding, membuatnya cocok bagi aplikasi dengan beban kerja tinggi dan volume data besar. Penggunaan NoSQL meliputi penanganan data semi-terstruktur (misalnya, HTML, XML, JSON), dengan optimasi yang menitikberatkan pada penyesuaian model data dan struktur fisik DBMS (Taipalus, 2024).

II.6.2 Jenis –Jenis Data

Data hadir dalam berbagai wujud dan format, tidak hanya sekadar kumpulan informasi yang seragam. Seperti yang telah didiskusikan sebelumnya, dalam ilmu basis data, data dapat dikategorikan menjadi beberapa jenis berdasarkan tingkat organisasi dan strukturnya. Data tersebut dikategorikan menjadi data terstruktur, data semi terstruktur data tidak terstruktur.

1. *Structured Data*

Data yang memiliki model yang terstruktur dengan baik dan mungkin mengikuti standar tertentu (Sarker, 2021). Ciri utamanya adalah setiap item dideskripsikan oleh kumpulan atribut yang sama, dan atribut-atribut ini memiliki kumpulan nilai yang diketahui (Ricci dkk., 2010).

2. *Semi-structured Data*

Data semi terstruktur seperti data terstruktur, namun memiliki organisasi melalui metadata, tag, atau penanda lain yang memisahkan elemen-elemen semantik. Data ini sering kali terdiri dari beberapa metadata generik yang mendeskripsikan karakteristik utama item, disertai dengan teks bebas atau elemen tidak terstruktur (Sarker, 2021).

3. *Unstructured Data*

Data yang tidak terstruktur adalah informasi yang tidak memiliki model data yang terdefinisi atau model yang sulit diproses langsung oleh komputer. Jenis data ini meliputi audio, video, dan teks dalam format bebas (Sarker, 2021).

II.7 *Incremental Learning*

Pembelajaran Kelas-Incremental mengacu pada proses pembelajaran dalam skenario dunia nyata di mana data terbatas dan data baru terus-menerus diberikan. Pendekatan ini dikategorikan ke dalam berbagai kelompok untuk mencapai keseimbangan antara kinerja, skalabilitas, efisiensi, dan kompleksitas dalam bidang pembelajaran mendalam (Tian dkk., 2024).

Sachi Nandan Mohanty, (2020) menjelaskan pada sistem rekomendasi, pembelajaran inkremental bukan hanya berfungsi sebagai fondasi pengembangan sistem *machine learning* yang responsif di lingkungan yang berubah-ubah, namun juga berperan sebagai elemen penting yang memungkinkan sistem rekomendasi memberikan layanan efektif kepada pengguna melalui informasi yang tepat sasaran dan terus diperbaharui, sekaligus mengatasi berbagai tantangan mendasar seperti keterbatasan data dan pergeseran preferensi pengguna.

II.8 *Data Streaming Process*

Pemrosesan aliran data mengacu pada eksekusi berkelanjutan dari tugas pemrosesan data pada aliran data yang berpotensi tidak terbatas. Proses ini mencakup pengumpulan, pemrosesan, dan analisis data secara langsung dari berbagai sumber, seperti jaringan sensor, pasar keuangan, dan streaming video. Dalam implementasinya, pemrosesan aliran data direpresentasikan sebagai grafik aliran data yang terdiri dari sumber data, pemroses data, dan tujuan akhir data (data sink). Hasil yang telah diproses kemudian dikirim ke aplikasi eksternal untuk dianalisis lebih lanjut (Qin dkk., 2019).

Dalam konteks sistem rekomendasi, prinsip pembelajaran inkremental, yang selaras dengan konsep pembelajaran aktif dan adaptasi berkelanjutan, sangat penting (Alfaifi, 2024). Pemrosesan aliran data memainkan peran penting dalam memungkinkan sistem rekomendasi untuk berfungsi secara efektif dengan cara-cara berikut:

1. Sebagai fondasi pengembangan sistem pembelajaran mesin yang responsif di lingkungan yang berubah-ubah (Alfaifi, 2024).
2. Elemen penting untuk layanan efektif melalui informasi yang tepat sasaran dan terus diperbarui (Ricci dkk., 2010).
3. Mengatasi tantangan mendasar seperti keterbatasan data dan pergeseran preferensi pengguna (Alfaifi, 2024).

II.9 *Penelitian terkait*

Nan dkk. (2022) mengusulkan kombinasi *data mining* dan *collaborative filtering* melalui proses *collaborative mining and filtering process* (CMFP) untuk memproses data pariwisata berskala besar. Studi ini menekankan efisiensi analisis perilaku pengguna dan data pengeluaran.

Javadian Sabet dkk. (2022) memperkenalkan *The Hybrid Offer Ranker* (THOR), yang menggabungkan beberapa algoritma *classification* (KNN, SVC, *Decision Tree*, *Random Forest*, *Logistic Regression*) untuk memperkirakan probabilitas

pembelian paket wisata. Sistem ini berhasil menangani *cold-start* menggunakan pendekatan klasterisasi (K-means, DBSCAN). Namun, penelitian ini lebih menekankan prediksi penjualan penawaran wisata.

Fararni dkk. (2021) menyajikan kerangka konseptual *hybrid recommender system* berbasis *big data* dan AI untuk merancang *trip planner* yang komprehensif. Walau arsitektur yang diusulkan dapat mengelola volume data wisata besar.

Shafqat dan Byun, (2020) mengembangkan *hierarchical LSTM* yang memprediksi lokasi wisata berikutnya dari riwayat perjalanan, kemudian menambahkan data konteks (cuaca, popularitas lokasi) melalui modul *feature importance* (XGBoost). Sistem menunjukkan peningkatan akurasi rekomendasi akurasi tertinggi yang dicapai oleh pendekatan *hierarchical LSTM* berada di kisaran 97,2%. Jika dibandingkan dengan model lain seperti GRU (96,4%) atau Bi-LSTM (94,2%), peningkatan akurasinya berada di rentang 0,8–3%.

Noorian Avval dan Harounabadi, (2023) pada penelitiannya mengusulkan *hybrid recommender system* yang memanfaatkan *topic modeling* dan *PrefixSpan algorithm* untuk rute wisata. Metode ini menggunakan *two-level clustering* (berbasis DBSCAN) untuk mengurangi waktu pencarian POI, serta menggabungkan *contextual data* dan analisis *asymmetric user similarity*. Hasilnya menunjukkan peningkatan akurasi dan efisiensi dengan peningkatan berada pada kisaran 5–8%.

Kumar dkk. (2024) berfokus pada *group recommendation system* (GRS) dengan mengeksplorasi karakteristik user-item dan rating kolaboratif. Mereka mengusulkan teknik baru untuk memprediksi rating individual dalam grup dan memperkenalkan fungsi agregasi (*Tendency-based Aggregation*) yang mampu meningkatkan kepuasan kelompok.

Shambour dkk. (2024) mengembangkan pendekatan *User-based Multi-Criteria Trust-aware Collaborative Filtering* (UMCTCF) yang menggabungkan *multi-criteria user preferences* dan *implicit trust network* untuk rekomendasi pariwisata. Pendekatan ini berhasil meningkatkan akurasi dan cakupan (*coverage*) rekomendasi, terutama dalam situasi data sparsity atau cold-start.

Puspasari (2022) penelitian ini melakukan penerapan berbagai model *machine learning* (antara lain KNN, SVM, *Decision Tree*, *Random Forest*, dan *Neural Network*) untuk merekomendasikan ruang pameran museum yang sesuai dengan profil dan minat pengunjung. Hasil eksperimen menunjukkan bahwa KNN memberikan kinerja terbaik dengan *cross-validation accuracy* mencapai 89,09% serta *F-Measure* sebesar 90,91%. Selain itu, uji kegunaan (*System Usability Scale*) untuk fitur rekomendasi pada aplikasi tur virtual mendapatkan skor rata-rata 85,83, yang menandakan tingkat penerimaan dan efisiensi yang cukup tinggi dalam membantu pengunjung menemukan pameran yang relevan.

Untuk memastikan destinasi wisata yang kurang populer mendapatkan kesempatan yang setara untuk direkomendasikan, penelitian ini menerapkan strategi mitigasi bias. Pendekatan ini secara khusus mengacu pada metode penanganan bias popularitas yang telah dikembangkan oleh (Yalcin & Bilge, 2021). Selain itu, konsep "*rekomendasi serendipity*" (Ricci dkk., 2010) juga diadopsi untuk memunculkan pilihan destinasi yang tak terduga namun relevan. Implementasi teknisnya akan melibatkan mekanisme penyesuaian bobot skor berdasarkan tingkat popularitas destinasi, sebuah teknik yang sejalan dengan gagasan dari (Kagie dkk., 2009).

Yoon dan Choi (2023) memperkenalkan sistem R2Tour yang memanfaatkan faktor konteks secara *real-time* (seperti kondisi cuaca dan jarak) beserta profil wisatawan (misalnya usia, gender, dan tipe perjalanan) untuk merekomendasikan destinasi wisata terdekat. Menggunakan data Jeju Island (Korea), R2Tour dilatih dengan beberapa model *machine learning* dan dievaluasi dengan akurasi mencapai 77,3%, *micro-F1* sebesar 0,773, serta *macro-F1* sebesar 0,415.

Berdasarkan penelitian sebelumnya, masih terdapat beberapa *gap* yang perlu diatasi, antara lain:

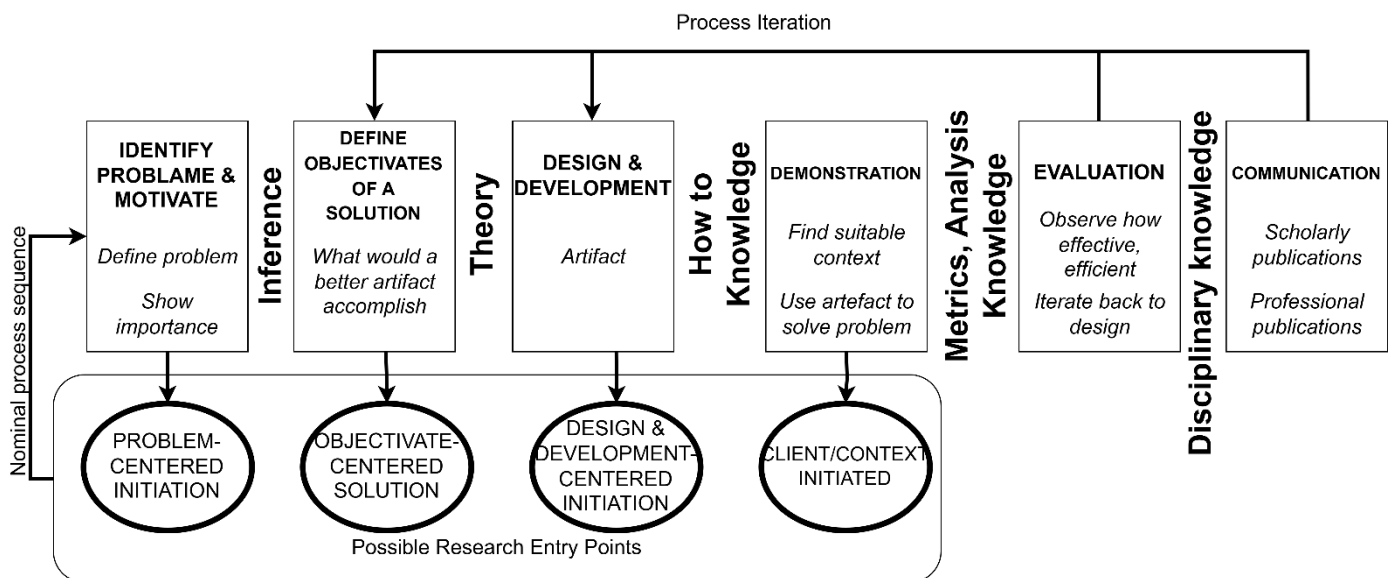
1. Integrasi data *real-time* (cuaca, kondisi lalu lintas, tren media sosial) yang belum optimal untuk meningkatkan relevansi dan adaptivitas rekomendasi.

2. Penanganan *cold-start* dan pembaruan model secara berkelanjutan (*incremental learning*) agar akurasi tidak menurun saat menghadapi pengguna atau destinasi baru.
3. Skalabilitas dalam menangani volume data pariwisata yang besar dan beragam, khususnya jika ingin menerapkan sistem rekomendasi adaptif secara *real-time*.

BAB III. Metodologi Penelitian

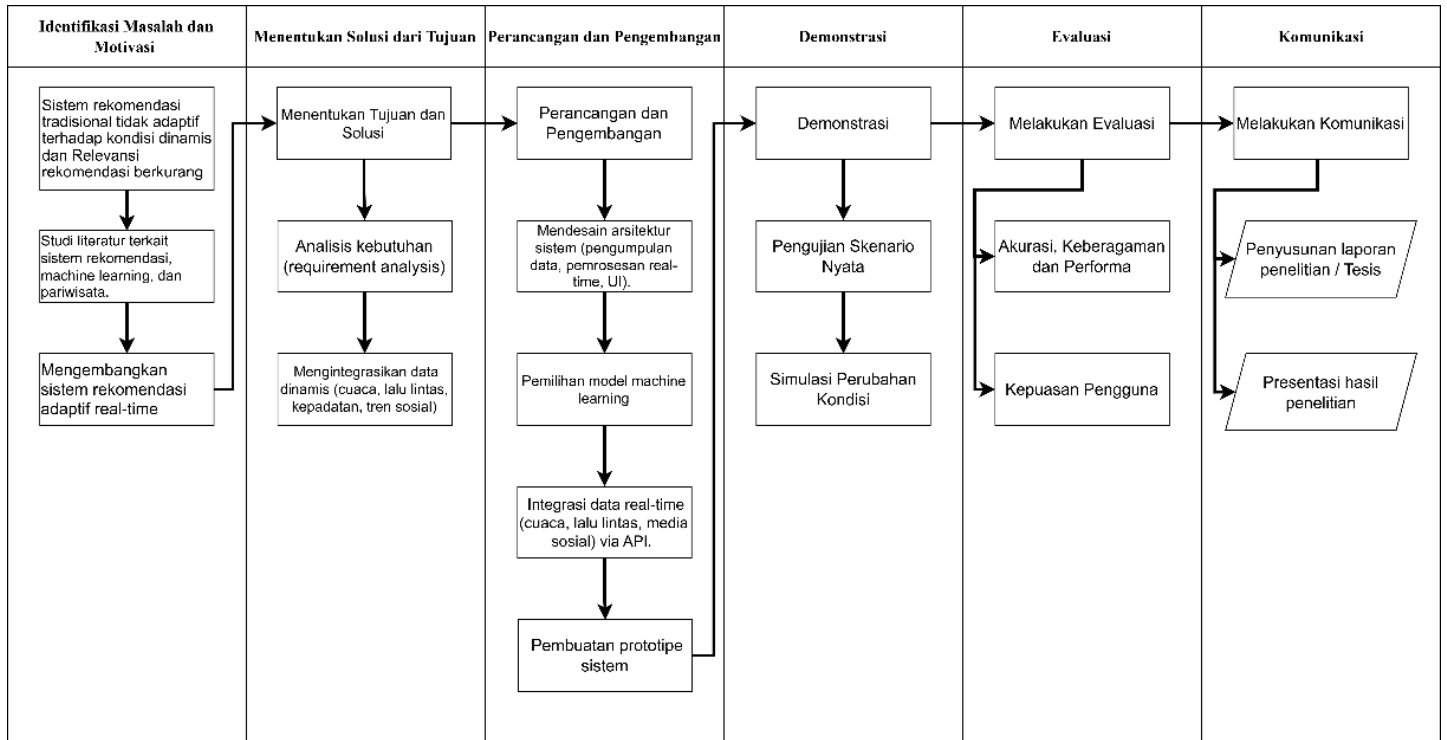
III.1 Metode Penelitian DSRM

Penelitian ini menggunakan metode DSRM (Design Science Research Methodology). DSRM adalah salah satu metode penelitian yang menyajikan setiap tahapan dengan mudah sehingga membantu dalam penelitian yang berhubungan dengan teknologi informasi dengan prinsip “pencarian solusi yang bertujuan” (Peffers dkk., 2007). Metode ini dapat digunakan untuk memahami dan melakukan evaluasi. DSRM terdiri dari enam tahapan mulai dari proses identifikasi terhadap masalah, menentukan solusi, melakukan perancangan dan pengembangan, melakukan demonstrasi, melakukan evaluasi, sampai dengan melakukan komunikasi seperti yang ditampilkan pada Gambar III. 1.



Gambar III.1 Kerangka Kerja DSRM berdasarkan Peffers dkk (2007)

Berdasarkan kerangka DSRM, setiap tahapan penelitian telah diuraikan metode pelaksanaannya secara rinci guna memastikan penelitian berjalan secara terukur dan terencana. DSRM dipilih karena memungkinkan iterasi desain-evaluasi untuk sistem adaptif. Misalnya, tahap 'demonstrasi' pada DSRM memfasilitasi uji coba prototipe dalam skenario perubahan cuaca mendadak. Berikut kerangka penelitian yang digunakan dalam penelitian ini sebagaimana pada Gambar III 2.



Gambar III.2 Kerangka Penelitian

III.2 Identifikasi Masalah dan Motivasi

Penelitian ini dimulai dengan tahap *Identifikasi Masalah dan Motivasi* dalam kerangka *Design Science Research Methodology (DSRM)*. Tujuan utama dari tahap ini adalah menyoroti keterbatasan sistem rekomendasi pariwisata yang ada saat ini. Permasalahan utama terletak pada ketidakmampuan sistem rekomendasi konvensional dalam beradaptasi terhadap kondisi yang dinamis, sehingga relevansi rekomendasi cenderung menurun. Sistem tersebut tidak mampu menyesuaikan secara *real-time* terhadap perubahan seperti variasi cuaca, kondisi lalu lintas, kepadatan pengunjung, dan tren di media sosial. Untuk mengidentifikasi celah ini secara sistematis, dilakukan studi literatur yang mendalam terhadap berbagai pendekatan sistem rekomendasi.

Motivasi dari penelitian ini adalah untuk mengembangkan sistem rekomendasi adaptif *real-time* yang mampu meningkatkan pengalaman wisatawan melalui penyediaan rekomendasi yang tepat waktu dan personal. Sistem ini diharapkan dapat mendukung pengambilan keputusan perjalanan yang lebih baik.

III.3 Menentukan Solusi dari Tujuan

Pada tahap *Menentukan Solusi dari Tujuan*, penelitian ini menitikberatkan pada perumusan visi sistem yang jelas serta analisis kebutuhan (*requirement analysis*) berdasarkan permasalahan yang telah teridentifikasi sebelumnya. Sasaran utamanya adalah membangun sistem rekomendasi adaptif *real-time* yang mampu memproses data dinamis dari berbagai sumber, seperti layanan cuaca, data lalu lintas, dan tren media sosial, guna menyajikan saran yang lebih relevan dan kontekstual bagi wisatawan.

Sistem ini juga dirancang untuk meningkatkan pengalaman pengguna melalui rekomendasi yang personal dan responsif terhadap perubahan kondisi, dengan tetap menjaga performa yang optimal. Dengan demikian, tujuan solusi yang dirumuskan menjadi dasar bagi proses perancangan dan evaluasi sistem yang efektif dan inovatif, dalam rangka menjawab tantangan adaptivitas *real-time* pada layanan rekomendasi pariwisata.

III.4 Perancangan dan Pengembangan

Tahap *Perancangan dan Pengembangan* merupakan proses pembangunan artefak teknologi berdasarkan tujuan yang telah dirumuskan. Langkah awal dilakukan dengan merancang arsitektur sistem yang terdiri atas tiga komponen utama: pengumpulan data, pemrosesan data secara *real-time*, dan antarmuka pengguna. Integrasi data dilakukan melalui pemanfaatan *Application Programming Interface* (API) dari berbagai penyedia layanan, seperti cuaca, lalu lintas, dan media sosial.

Selanjutnya, dikembangkan model *machine learning* adaptif yang mampu mempelajari pola secara berkelanjutan. Algoritma rekomendasi dirancang untuk menggabungkan preferensi pengguna dengan kondisi lingkungan terkini. Antarmuka pengguna dirancang secara intuitif agar mendukung interaksi yang mudah dan efisien. Seluruh komponen ini disatukan dalam bentuk prototipe sistem yang siap untuk diuji pada tahap demonstrasi.

III.5 Demonstrasi

Tahap *Demonstrasi* bertujuan untuk menunjukkan bagaimana artefak sistem dapat digunakan untuk menyelesaikan permasalahan yang telah dirumuskan. Demonstrasi dilakukan melalui simulasi skenario nyata, seperti perubahan cuaca mendadak, kemacetan lalu lintas, dan lonjakan kepadatan pengunjung. Tujuannya adalah untuk menilai kemampuan sistem dalam menyesuaikan rekomendasi secara adaptif.

Selain itu, prototipe diuji coba oleh sekelompok pengguna terbatas untuk memperoleh umpan balik awal terkait kegunaan, kenyamanan, dan relevansi hasil rekomendasi. Hasil demonstrasi menjadi validasi awal terhadap efektivitas sistem serta bahan pertimbangan untuk penyempurnaan pada tahap evaluasi.

III.6 Evaluasi

Tahap *Evaluasi* dilakukan untuk menilai sejauh mana artefak sistem rekomendasi memenuhi tujuan yang telah ditetapkan. Evaluasi dilakukan menggunakan metrik kuantitatif seperti akurasi, keberagaman rekomendasi, kecepatan respons, dan stabilitas pemrosesan data. Evaluasi kualitatif dilakukan melalui survei atau wawancara terhadap pengguna untuk mengukur persepsi terhadap kegunaan dan kepuasan penggunaan sistem.

Sebagai pembanding, sistem ini dievaluasi terhadap pendekatan rekomendasi statis guna menyoroti keunggulan adaptivitas *real-time*. Apabila hasil evaluasi menunjukkan adanya kekurangan, maka dilakukan iterasi kembali ke tahap *Perancangan dan Pengembangan* untuk perbaikan.

III.7 Komunikasi

Tahap *Komunikasi* merupakan fase akhir dalam DSRM yang berfokus pada diseminasi hasil penelitian kepada komunitas akademik dan praktisi. Komunikasi dilakukan melalui penyusunan tesis yang mendokumentasikan seluruh proses, temuan, dan artefak sistem secara komprehensif.

BAB IV. Analisis Dan Desain Awal

IV.1 Analisis Kebutuhan Sistem

Sistem ini bertujuan untuk menyediakan rekomendasi pariwisata adaptif *real-time* yang disesuaikan dengan preferensi pengguna dan kondisi aktual di lapangan. Berdasarkan studi literatur, kebutuhan sistem dapat diklasifikasikan menjadi:

IV.1.1 Kebutuhan Fungsional

Sistem harus memiliki fungsi utama sebagai berikut:

1. Sistem harus dapat mengakses dan mengumpulkan data cuaca, kondisi lalu lintas, kepadatan pengunjung, serta tren media sosial secara berkala melalui API atau sumber data lain.
2. Sistem harus memproses data ini untuk memberikan rekomendasi personal dan adaptif.
3. Sistem harus memiliki antarmuka pengguna yang mudah digunakan dan aman.
4. Sistem harus menghasilkan rekomendasi personal berdasarkan preferensi pengguna dan kondisi *real-time*.
5. Sistem harus terus memantau dan menyesuaikan rekomendasi berdasarkan perubahan, seperti lonjakan kepadatan atau perubahan tren sosial.

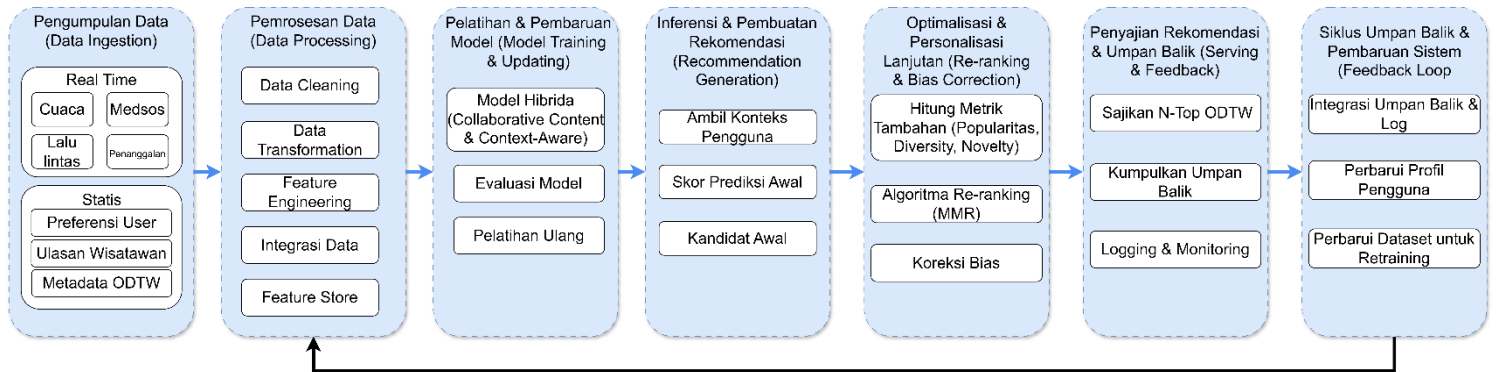
IV.1.2 Kebutuhan Non-Fungsional

Sistem harus memenuhi beberapa kriteria non-fungsional berikut:

1. Sistem harus dapat memproses data *real-time* dan menampilkan rekomendasi dengan latensi rendah.
2. Akurasi dan keberagaman tinggi dalam memberikan rekomendasi.
3. Sistem harus melindungi data pengguna, seperti preferensi dan riwayat perjalanan, sesuai regulasi perlindungan data.
4. Sistem harus andal dengan downtime minimal, tersedia 24/7 untuk mendukung wisatawan kapan saja.

IV.2 Model Konseptual Sistem

Model konseptual sistem rekomendasi adaptif real-time yang diusulkan dalam penelitian ini dirancang dengan arsitektur pipeline data komprehensif yang terdiri dari tujuh tahapan utama, sebagaimana diilustrasikan pada Gambar IV.1.



Gambar IV.1 Arsitektur sistem yang akan dikembangkan

Arsitektur ini, dirancang untuk memproses data secara sistematis, mulai dari akuisisi hingga penyajian rekomendasi dan pemanfaatan umpan balik untuk pembelajaran berkelanjutan. Setiap tahapan melibatkan metode dan teknologi spesifik untuk memastikan efisiensi dan efektivitas sistem. Berikut penjelasan singkat setiap tahap pada diagram alur sistem rekomendasi:

1. Tahap 1: Pengumpulan Data (Data Ingestion)
 - a. Mengumpulkan data real-time (cuaca, lalu lintas, tren sosial media, acara) melalui API publik menggunakan library Python requests dan dijadwalkan dengan APScheduler.
 - b. Menyimpan data mentah real-time pada staging area berupa local file system dalam format JSON.
 - c. Mengekstraksi data statis/historis (profil pengguna, metadata ODTW, ulasan) dari basis data relasional lokal MySQL atau dari berkas JSON menggunakan pandas.
 - d. Menyimpan data historis di local file system.
2. Tahap 2: Pemrosesan Data (Data Processing)
 - a. Melakukan pembersihan data (data cleaning) untuk menangani nilai hilang, duplikasi, dan outlier menggunakan pandas dan scikit-learn.
 - b. Melakukan transformasi data.

- c. Melakukan rekayasa fitur (feature engineering).
 - d. Mengintegrasikan seluruh fitur menjadi dataset terpadu menggunakan *pandas*.
 - e. Menyimpan *DataFrame* fitur yang sudah diproses sebagai berkas *Parquet* di *local file system*.
3. Tahap 3: Pelatihan dan Pembaruan Model (Model Training & Updating)
- a. Melatih model *machine learning* (ML) menggunakan Model Hibrida Collaborative Filtering dan model berbasis konten yang memanfaatkan fitur ODTW dan fitur *real-time*.
 - b. Mengevaluasi model menggunakan metrik standar (MAE, RMSE, Precision, Recall, F1-score, NDCG).
 - c. Menerapkan pelatihan ulang model secara periodik.
4. Tahap 4: Inferensi dan Pembuatan Rekomendasi (Recommendation Generation)
- a. Mengambil profil pengguna dan data kontekstual *real-time* yang relevan saat ada permintaan.
 - b. Menggunakan model ML yang telah dilatih untuk menghasilkan skor prediksi relevansi untuk setiap destinasi.
 - c. Memilih sejumlah destinasi dengan skor tertinggi (Top-K) sebagai kandidat rekomendasi awal melalui pengurutan skor.
5. Tahap 5: Optimalisasi dan Personalisasi Lanjutan (Re-ranking & Bias Correction)
- a. Menghitung metrik tambahan untuk kandidat (*popularitas, kebaruan*)
 - b. Menerapkan algoritma re-ranking dengan implementasi manual Maximal Marginal Relevance (MMR) untuk meningkatkan keberagaman.
 - c. Merancang dan mengimplementasikan heuristik atau aturan spesifik untuk koreksi bias popularitas guna mengangkat destinasi kurang populer yang relevan.
6. Tahap 6: Penyajian Rekomendasi dan Umpan Balik (Recommendation Serving & Feedback)
- a. Menyajikan N-Top ODTW yang telah dioptimalkan kepada pengguna melalui *endpoint* API sederhana yang dibuat dengan *Flask*.

- b. Mengumpulkan umpan balik pengguna.
 - c. Mencatat aktivitas sistem
- 7. Tahap 7: Siklus Umpan Balik dan Pembaruan Sistem (*Feedback Loop*)
 - a. Memproses data umpan balik yang terkumpul.
 - b. Menggunakan informasi umpan balik untuk memperbarui profil pengguna dan menjadi input untuk proses pelatihan ulang model ML.

IV.3 Rencana Jadwal Penelitian

Pada penelitian ini rencana jadwal yang akan dilaksanakan, seperti yang ditunjukkan pada tabel IV.1

Table IV.1 Rencana Jadwal Penelitaian

Kegiatan	April				Mei				Juni				Juli				Agustus				September				Oktober				November				Desember			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4				
Finalisasi Pendahuluan																																				
Finalisasi Tinjaun Pustaka																																				
Finalisasi Metodologi																																				
Finalisasi Bagian Analisis																																				
Finalisasi Bagian Desain																																				
Finalisasi Evaluasi dan Kesimpulan																																				
Persiapan Sidang Magister																																				

DAFTAR PUSTAKA

- Achmad, F., Prambudia, Y., & Rumanti, A. A. (2023). Improving Tourism Industry Performance through Support System Facilities and Stakeholders: The Role of Environmental Dynamism. *Sustainability (Switzerland)*, 15(5).
<https://doi.org/10.3390/su15054103>
- Akhadam, A., Kbibchi, O., Mekouar, L., & Iraqi, Y. (2025). A Comparative Evaluation of Recommender Systems Tools. *IEEE Access*.
<https://doi.org/10.1109/ACCESS.2025.3541014>
- Alfaifi, Y. H. (2024). Recommender Systems Applications: Data Sources, Features, and Challenges. Dalam *Information (Switzerland)* (Vol. 15, Nomor 10). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/info15100660>
- Alotaibi, O., Pardede, E., & Tomy, S. (2023). Cleaning Big Data Streams: A Systematic Literature Review. Dalam *Technologies* (Vol. 11, Nomor 4). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/technologies11040101>
- Barykin, S. E., de la Poza, E., Khalid, B., Kapustina, I. V., Kalinina, O. V., & Iqbal, K. M. J. (2021). Tourism industry: Digital transformation. Dalam *Handbook of Research on Future Opportunities for Technology Management Education* (hlm. 414–434). IGI Global. <https://doi.org/10.4018/978-1-7998-8327-2.ch025>
- Chalkiadakis, G., Ziogas, I., Koutsmanis, M., Streviniotis, E., Panagiotakis, C., & Papadakis, H. (2023). A Novel Hybrid Recommender System for the Tourism Domain. *Algorithms*, 16(4). <https://doi.org/10.3390/a16040215>
- Choi, I. Y., Ryu, Y. U., & Kim, J. K. (2021). A recommender system based on personal constraints for smart tourism city. *Asia Pacific Journal of Tourism Research*, 26(4), 440–453. <https://doi.org/10.1080/10941665.2019.1592765>
- DISPARBUDPORA. (2024). *DATA KUNJUNGAN OBJEK WISATA TAHUN 2023*.
- Falk, K. (2019). *Practical Recommender Systems*.
- Fararni, K. Al, Nafis, F., Aghoutane, B., Yahyaouy, A., Riffi, J., & Sabri, A. (2021). Hybrid recommender system for tourism based on big data and AI: A conceptual framework. *Big Data Mining and Analytics*, 4(1), 47–55.
<https://doi.org/10.26599/BDMA.2020.9020015>
- Fragkoulis, M., Carbone, P., Kalavri, V., & Katsifodimos, A. (2024). A survey on the evolution of stream processing systems. *VLDB Journal*, 33(2), 507–541.
<https://doi.org/10.1007/s00778-023-00819-8>
- Huang, S., Qin, Y., Zhang, X., Tu, Y., Li, Z., & Cui, B. (2023). Survey on performance optimization for database systems. Dalam *Science China Information Sciences* (Vol. 66, Nomor 2). Science Press (China). <https://doi.org/10.1007/s11432-021-3578-6>
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). *Machine learning and deep learning*. <https://doi.org/10.1007/s12525-021-00475-2/Published>

- Javadian Sabet, A., Shekari, M., Guan, C., Rossi, M., Schreiber, F., & Tanca, L. (2022). THOR: A Hybrid Recommender System for the Personalized Travel Experience. *Big Data and Cognitive Computing*, 6(4). <https://doi.org/10.3390/bdcc6040131>
- Jie Tang, Jimeng Sun, Chi Wang, & Zi Yang. (2013). *Social Influence Analysis in Large-scale Networks*. ACM Digital Library.
- Kagie, M., Van Wezel, M., & Groenen, P. J. F. (2009). *Determination of Attribute Weights for Recommender Systems Based on Product Popularity*. <https://www.researchgate.net/publication/46433887>
- Kamal, N., Sarker, F., Rahman, A., Hossain, S., & Mamun, K. A. (2024). Recommender System in Academic Choices of Higher Education: A Systematic Review. *IEEE Access*, 12, 35475–35501. <https://doi.org/10.1109/ACCESS.2024.3368058>
- Kumar, J., Patra, B. K., Sahoo, B., & Babu, K. S. (2024). Group recommendation exploiting characteristics of user-item and collaborative rating of users. *Multimedia Tools and Applications*, 83(10), 29289–29309. <https://doi.org/10.1007/s11042-023-16799-4>
- Nan, X., Kayo Kanato, & Wang, X. (2022). Design and Implementation of a Personalized Tourism Recommendation System Based on the Data Mining and Collaborative Filtering Algorithm. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/1424097>
- Noorian, A. (2024). A personalized context and sequence aware point of interest recommendation. *Multimedia Tools and Applications*, 83(32), 77565–77594. <https://doi.org/10.1007/s11042-024-18522-3>
- Noorian Avval, A. A., & Harounabadi, A. (2023). A hybrid recommender system using topic modeling and prefixspan algorithm in social media. *Complex and Intelligent Systems*, 9(4), 4457–4482. <https://doi.org/10.1007/s40747-022-00958-5>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Dalam *Journal of Management Information Systems* (Vol. 24, Nomor 8). <http://www.tuunanen.fi>
- Pencarelli, T. (2020). The digital revolution in the travel and tourism industry. *Information Technology and Tourism*, 22(3), 455–476. <https://doi.org/10.1007/s40558-019-00160-3>
- Pugliese, R., Regondi, S., & Marini, R. (2021). Machine learning-based approach: Global trends, research directions, and regulatory standpoints. Dalam *Data Science and Management* (Vol. 4, hlm. 19–29). KeAi Communications Co. <https://doi.org/10.1016/j.dsm.2021.12.002>
- Puspasari, S. (2022). Machine Learning for Exhibition Recommendation in a Museum's Virtual Tour Application. Dalam *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 13, Nomor 4). www.ijacsa.thesai.org

- Qin, C., Eichelberger, H., & Schmid, K. (2019). Enactment of adaptation in data stream processing with latency implications—A systematic literature review. *Information and Software Technology*, 111, 1–21. <https://doi.org/10.1016/j.infsof.2019.03.006>
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2010). *Recommender Systems Handbook*.
- Rumanti, A. A., Sunaryo, I., Wiratmadja, I. I., & Irianto, D. (2023). Cleaner Production for Small and Medium Enterprises: An Open Innovation Perspective. *IEEE Transactions on Engineering Management*, 70(7), 2355–2368. <https://doi.org/10.1109/TEM.2020.3015048>
- Sachi Nandan Mohanty, J. M. C. S. J. A. A. E. P. G. (2020). *Recommender System with Machine Learning and Artificial Intelligence*.
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. Dalam *SN Computer Science* (Vol. 2, Nomor 3). Springer. <https://doi.org/10.1007/s42979-021-00592-x>
- Shafqat, W., & Byun, Y. C. (2020). A context-aware location recommendation system for tourists using hierarchical LSTM model. *Sustainability (Switzerland)*, 12(10). <https://doi.org/10.3390/su12104107>
- Shambour, Q. Y., Abualhaj, M. M., Abu-Shareha, A. A., & Kharma, Q. M. (2024). PERSONALIZED TOURISM RECOMMENDATIONS: LEVERAGING USER PREFERENCES AND TRUST NETWORK. *Interdisciplinary Journal of Information, Knowledge, and Management*, 19. <https://doi.org/10.28945/5329>
- Shuvo, M. I. M., & Islam, M. J. (2024). The digital transformation of tourism: a study of tourist behaviour and preferences in the age of technology in Bangladesh. *Research in Hospitality Management*, 14(3), 236–244. <https://doi.org/10.1080/22243534.2024.2419366>
- Singh, S. K., Tiwari, A. K., & Paliwal, H. K. (2023). A state-of-the-art review on the utilization of machine learning in nanofluids, solar energy generation, and the prognosis of solar power. Dalam *Engineering Analysis with Boundary Elements* (Vol. 155, hlm. 62–86). Elsevier Ltd. <https://doi.org/10.1016/j.enganabound.2023.06.003>
- Solano-Barliza, A., Arregocés-Julio, I., Aarón-Gonzalvez, M., Zamora-Musa, R., De-La-Hoz-Franco, E., Escorcia-Gutierrez, J., & Acosta-Coll, M. (2024). Recommender systems applied to the tourism industry: a literature review. Dalam *Cogent Business and Management* (Vol. 11, Nomor 1). Cogent OA. <https://doi.org/10.1080/23311975.2024.2367088>
- Statista Research Department. (2025, Februari 21). *Artificial intelligence (AI) use in travel and tourism - statistics & facts*.
- Suhaim, A. Bin, & Berri, J. (2021). Context-Aware Recommender Systems for Social Networks: Review, Challenges and Opportunities. *IEEE Access*, 9, 57440–57463. <https://doi.org/10.1109/ACCESS.2021.3072165>

- Taipalus, T. (2024). Database management system performance comparisons: A systematic literature review. *Journal of Systems and Software*, 208. <https://doi.org/10.1016/j.jss.2023.111872>
- Tian, S., Li, L., Li, W., Ran, H., Ning, X., & Tiwari, P. (2024). A survey on few-shot class-incremental learning. *Neural Networks*, 169, 307–324. <https://doi.org/10.1016/j.neunet.2023.10.039>
- World Tourism Organization. (2024). *International Tourism Highlights, 2024 Edition*. UN Tourism. <https://doi.org/10.18111/9789284425808>
- Yalcin, E., & Bilge, A. (2021). Investigating and counteracting popularity bias in group recommendations. *Information Processing and Management*, 58(5). <https://doi.org/10.1016/j.ipm.2021.102608>
- Yazici, İ., Shayea, I., & Din, J. (2023). A survey of applications of artificial intelligence and machine learning in future mobile networks-enabled systems. Dalam *Engineering Science and Technology, an International Journal* (Vol. 44). Elsevier B.V. <https://doi.org/10.1016/j.jestch.2023.101455>
- Yoon, J. H., & Choi, C. (2023). Real-Time Context-Aware Recommendation System for Tourism. *Sensors*, 23(7). <https://doi.org/10.3390/s23073679>