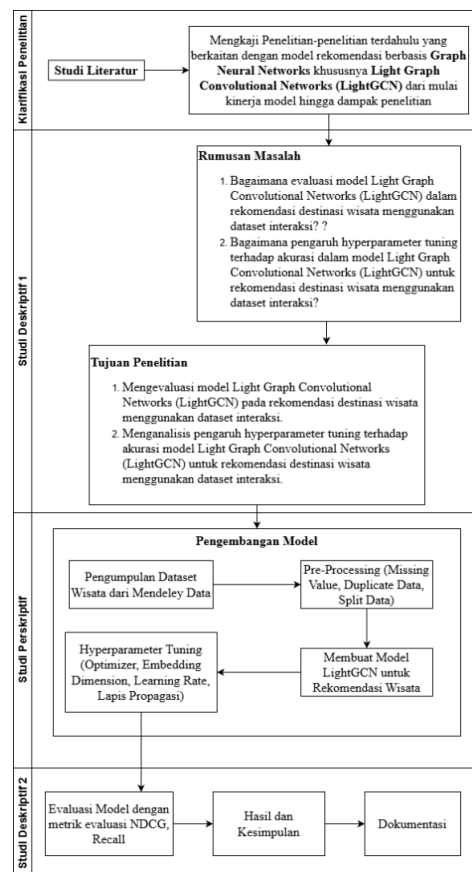


BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Metodologi penelitian yang digunakan adalah *Design Research Methodology* (DRM), yang dirancang untuk mendukung proses penelitian berbasis desain dengan pendekatan sistematis. Metode DRM ini cocok untuk riset yang bertujuan Praktis maupun teoritis, sementara model lain yaitu DSRM lebih mengarah pada perancangan solusi praktis untuk kasus nyata (McKenney & Handley, 2020). DRM terdiri dari empat tahap utama, yaitu *Research Clarification* (RC), *Descriptive Study I* (DS-I), *Prescriptive Study* (PS), dan *Descriptive Study II* (DS-II).



Gambar 3. 1 Diagram DRM Penelitian

3.2 Klarifikasi Penelitian

Tahap pertama pada metode DRM adalah Klarifikasi Penelitian, dilakukan untuk merangkum tujuan penelitian menurut latar belakang permasalahan sehingga dapat mengarahkan landasan penelitian. Pada tahap ini, dilakukan studi literatur terhadap berbagai penelitian sebelumnya yang membahas sistem rekomendasi dan mengapa model *Graph Neural Networks* (GNN) LightGCN memiliki keunggulan dibanding model rekomendasi lainnya untuk dikaji lebih lanjut. Referensi dalam merumuskan landasan penelitian diperoleh dari berbagai jurnal untuk memperoleh pengetahuan lebih mendalam mengenai model rekomendasi LightGCN. Dengan demikian, pada tahap ini ditetapkan bahwa penelitian akan menitik beratkan pada implementasi dan evaluasi model LightGCN dalam rekomendasi destinasi wisata lalu melakukan analisis terhadap pengaruh hyperparameter tuning.

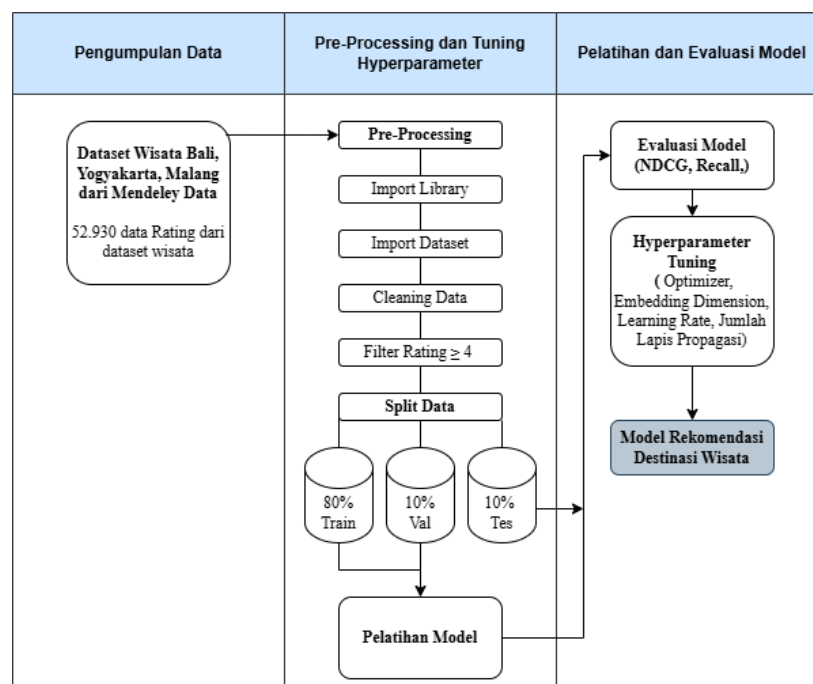
3.4 Studi Perskriptif

Setelah tahap Studi Deskriptif 1 selesai, penelitian memasuki tahap Studi Perskriptif yang berfokus pada pengembangan model untuk mengatasi permasalahan yang telah diidentifikasi. Dalam konteks penelitian ini, solusi yang dikembangkan adalah implementasi model sistem rekomendasi destinasi wisata dengan model *Graph Neural Networks* (GNN) LightGCN. Tahap ini mencakup perancangan arsitektur model, persiapan dataset interaksi antara pengguna dan destinasi wisata, serta proses pengembangan model LightGCN. Setelah arsitektur model dirumuskan, dataset interaksi pengguna dan wisata akan diproses melalui tahapan *preprocessing*, yang meliputi pembersihan data, normalisasi, dan pembagian dataset menjadi data latih, data validasi dan data uji. Langkah selanjutnya adalah pengembangan model, di mana algoritma LightGCN diterapkan pada struktur graf interaksi untuk memperkuat representasi hubungan antar pengguna dan destinasi. Lalu model yang telah dikembangkan akan dievaluasi menggunakan metrik NDCG, Recall. Proses ini diharapkan menghasilkan model rekomendasi yang mampu memberikan saran destinasi wisata yang lebih akurat dan relevan.

3.4.1 Pengembangan Model

Dalam tahap pengembangan model sistem rekomendasi destinasi wisata tentu terdapat beberapa tahapan di dalamnya yang menjadi *track of the research* dari proses pengembangan model ini. Model yang akan digunakan dalam penelitian ini adalah LightGCN yang dimana model ini memiliki tujuan untuk menangkap sinyal kolaboratif antara pengguna dan destinasi wisata secara efektif di dalam graf interaksi sehingga diharapkan dapat menghasilkan rekomendasi wisata yang relevan dan akurat.

Pengembangan model dibagi menjadi beberapa tahap yakni pengumpulan data, kemudian melalui tahap *preprocessing* dan *tuning hyperparameter* untuk mencari nilai konfigurasi parameter terbaik, Setelah melalui tahap tuning dan evaluasi, model yang telah memperoleh performa terbaik diimplementasikan untuk menghasilkan rekomendasi destinasi wisata. Implementasi ini nantinya akan dievaluasi menggunakan data pengujian untuk melihat keefektifan model dalam memberikan rekomendasi yang relevan dan akurat. Dibawah ini adalah Gambar 3.2 alur dari proses pengembangan model.



Gambar 3. 2 Prosedur Pengembangan Model Rekomendasi Wisata

Frahari Perdana Putra, 2025

OPTIMASI MODEL LIGHT GRAPH CONVOLUTIONAL NETWORK PADA REKOMENDASI DESTINASI WISATA BERDASARKAN RIWAYAT INTERAKSI PENGGUNA

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

3.4.2 Pengumpulan Data

Pada tahap pengumpulan data, penelitian ini menggunakan Dataset yang terpublikasi pada jurnal *Data in Brief*. Dataset ini mengolah data transaksi historis antara pengguna dengan destinasi wisata di 3 daerah wisata di Indonesia, yaitu Bali, Yogyakarta dan Malang. Dataset ini mencakup jenis data destinasi wisata yang dibutuhkan khususnya untuk sistem rekomendasi berbasis kolaboratif filtering yang membutuhkan data interaksi seperti *rating* yang terdapat pada data transaksi dengan ukuran sebanyak 52.930 data *rating*, selain itu dataset ini juga memiliki data pengguna serta data item atau dalam penelitian ini disebut destinasi wisata.

Data-data tersebut seluruhnya dapat diakses dari *Mendeley Data* dan bersifat *open source* sehingga dapat diunduh untuk keperluan penelitian model rekomendasi destinasi wisata.

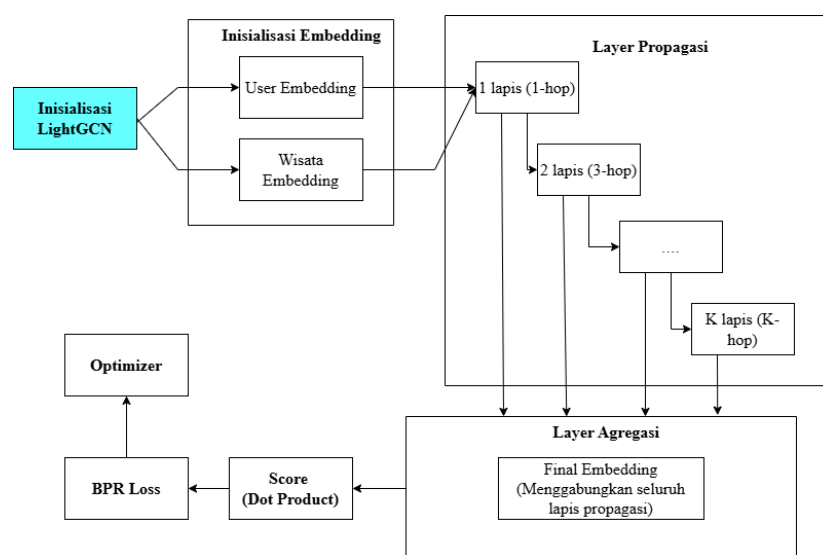
3.4.3 Preprocessing

Demi memastikan kualitas dan konsistensi data yang akan digunakan dalam membangun model sistem rekomendasi wisata, dilakukan tahapan pemrosesan data. Langkah ini diawali dengan memuat dataset yang terdiri dari 3 jenis data transaksi, data pengguna dan data item atau wisata. Setelah dimuat dataset ini kemudian dibersihkan dengan menghapus *missing values* dan juga duplikasi untuk memastikan data dalam kondisi layak digunakan. Bersamaan dengan tahap pembersihan data, dilakukan eksplorasi data untuk melihat karakteristik dataset dan konsistensi data seperti memastikan bahwa UserID pada data transaksi konsisten dengan pada data pengguna, juga AttractionID pada data transaksi konsisten dengan AttractionID pada data wisata.

Selanjutnya dilakukan pembagian proporsi data menjadi tiga jenis set data yaitu data latih (80%), data validasi (10%) dan data uji (10%), yang merujuk pada penelitian (Sachdeva dkk., 2022) yang menggunakan pembagian data tersebut pada berbagai macam dataset untuk model rekomendasi berbasis graf.

3.4.4 Arsitektur *Light Graph Convolutional Networks*

Selanjutnya tahapan yang dilakukan dalam penelitian ini adalah membangun skema arsitektur model dengan metode *Graph Neural Networks* (GNN) LightGCN, model ini dibangun dengan memanfaatkan struktur *graph* yang merepresentasikan interaksi antara pengguna dan wisata, serta menggunakan propagasi pesan linear pada graf untuk mempelajari representasi pengguna dan wisata.



Gambar 3. 3 Arsitektur proses pengembangan LightGCN

Berdasarkan gambar arsitektur 3.3 tersebut yang menunjukkan bagaimana model dikembangkan dalam penelitian ini, berikut ini penjelasan rinci mengenai skema arsitektur *Graph Neural Networks* (GNN) LightGCN:

1. Inisialisasi LightGCN

- Tahapan Inisialisasi LightGCN dalam penelitian ini merujuk pada proses pembuatan *instance* dari kelas LightGCN yang merepresentasikan model yang akan digunakan untuk vektor representasi pengguna dan item, serta *sparse adjacency matrix* yang telah dibangun untuk merepresentasikan koneksi antar pengguna dan item. Parameter-parameter ini disimpan sebagai atribut dalam

objek model, menyiapkan kerangka kerja yang diperlukan untuk mendefinisikan dan menjalankan operasi propagasi graf selanjutnya.

2. Inisialisasi *Embedding*

- Setiap pengguna dan setiap wisata dalam dataset diberikan representasi vektor awal yang disebut *embedding*. Setiap pengguna memiliki vektor *embedding* uniknya sendiri, begitu juga setiap wisata. *Embedding* awal ini diinisialisasi secara acak menggunakan inisialisasi *Xavier uniform* seperti pada Penelitian (G. Chen dkk., 2025), dan akan diperbarui selama proses pelatihan. *Embedding* ini menangkap fitur laten awal dari pengguna dan wisata.

3. Layer Propagasi

- Tahap ini adalah inti dari model rekomendasi berbasis graf, pada setiap lapisan, informasi dari *node* tetangga disebarkan dan digabungkan untuk memperbarui *embedding node* saat ini.
- Graf yang digunakan merepresentasikan interaksi Pengguna dan wisata, di mana pengguna dan wisata adalah *node*, dan interaksi adalah *edge* antar *node*.
- 1-hop : Pada lapisan pertama, *embedding* pengguna diperbarui dengan menggabungkan informasi dari wisata yang berinteraksi langsung dengannya, dan *embedding* wisata diperbarui dengan menggabungkan informasi dari pengguna yang berinteraksi dengannya. Ini menangkap interaksi langsung satu lompatan.
- 2-hop : Pada lapisan kedua, informasi disebarkan lebih jauh. *Embedding* pengguna kini menggabungkan informasi dari wisata yang berinteraksi dengannya, dan melalui wisata tersebut, menggabungkan informasi dari pengguna lain yang berinteraksi dengan wisata yang sama. Demikian pula untuk wisata. Ini menangkap hubungan tidak langsung, seperti pengguna yang memiliki preferensi serupa karena mereka berinteraksi dengan wisata yang sama.

- Proses ini berlanjut hingga K lapisan. Semakin banyak lapisan, semakin jauh informasi disebarkan dalam graf, memungkinkan model menangkap hubungan tidak langsung yang lebih kompleks antara pengguna dan item.
- LightGCN menggunakan *linear graph convolution* di setiap lapisan, yang berarti operasi agregasi dan transformasi pada setiap *node* hanya akan melibatkan penjumlahan sederhana dari *embedding* tetangga yang telah dinormalisasi. Ini berbeda dari GCN tradisional yang menggunakan transformasi linier dan fungsi aktivasi non-linear di setiap layer.

4. Layer Agregasi

- Setelah propagasi melalui K lapisan, setiap pengguna dan item kini memiliki sekumpulan *embedding* dari setiap lapisan termasuk *embedding* awal dari inisialisasi.
- Pada tahap agregasi, *embedding* dari semua lapisan untuk pengguna yang sama digabungkan untuk membentuk *embedding final* pengguna tersebut, hal yang sama juga dilakukan untuk wisata.
- metode penggabungan dilakukan dengan mengambil rata-rata dari *embedding* di semua lapisan. *Embedding* final ini dianggap sebagai representasi pengguna dan item yang paling kaya informasi setelah mempertimbangkan struktur graf interaksi.

5. Score

- Setelah mendapatkan *embedding final* untuk semua pengguna dan item, model menghitung skor prediksi untuk setiap pasangan pengguna dan wisata yang ingin direkomendasikan.
- Metode yang digunakan adalah *Dot Product* antara *embedding final* pengguna dan *embedding final* wisata seperti pada penelitian Hedkk., (2020). Skor hasil *dot product* ini merepresentasikan seberapa cocok pengguna tersebut dengan item tersebut. Skor yang lebih

tinggi menunjukkan kemungkinan bahwa pengguna akan tertarik dengan wisata tersebut.

6. BPR Loss

- Model ini akan menggunakan BPR loss seperti merujuk pada penelitian He dkk., (2020) fungsi kerugian yang digunakan untuk melatih model dengan data umpan balik implisit seperti data transaksi di mana hanya interaksi positif yang diketahui.
- *Loss Binary Pairwise Ranking* beroperasi berdasarkan pasangan triplet yaitu pengguna, wisata positif yaitu wisata yang berinteraksi dengan pengguna, dan wisata negatif yaitu wisata yang tidak berinteraksi dengan pengguna.
- Tujuan dari loss BPR adalah untuk memaksimalkan skor prediksi dari item positif relatif terhadap skor prediksi dari item negatif untuk pengguna yang sama. Selama pelatihan, model menyesuaikan bobot untuk mencapai ini.

7. Optimizer

- Optimizer merupakan algoritma kunci yang digunakan untuk memperbarui parameter model, khususnya embedding pengguna dan wisata, selama proses pelatihan. Berdasarkan gradien dari fungsi BPR Loss yang dihitung melalui backpropagation, optimizer menyesuaikan nilai embedding secara iteratif dengan tujuan meminimalkan loss.
- Berdasarkan gradien dari fungsi BPR Loss yang dihitung melalui backpropagation, *Optimizer* menyesuaikan nilai *embedding* secara iteratif dengan tujuan meminimalkan loss.
- Optimizer yang berbeda beda akan diimplementasikan lalu diuji pada penelitian ini, diantaranya adalah Adam, RMSprop, SGD, Adagrad dan Nadam seperti merujuk pada penelitian (Anggara dkk., 2023) .

3.4.5 Pengembangan Model

Dalam penelitian ini, pelatihan model dimulai dengan pembuatan *instance* model LightGCN dengan konfigurasi parameter *default* seperti merujuk pada penelitian He dkk., (2020) dengan konfigurasi *embedding dimension* 64, *learning rate* 0.001, dan jumlah lapis propagasi diuji dari 1 sampai 4 yang digunakan diseluruh dataset yang diuji, proses pelatihan model akhir ini dilakukan dalam sebuah *loop* iteratif yang berjalan selama jumlah epoch maksimum yang ditentukan dengan 50 epoch atau hingga mekanisme *early stopping* diaktifkan, pada setiap epoch, data training diproses dalam batch-batch. Untuk setiap batch, dilakukan negative sampling untuk menghasilkan pasangan pengguna-item positif yaitu interaksi yang ada dan pengguna-item negatif yaitu interaksi yang tidak ada. Model kemudian menghitung skor prediksi untuk pasangan positif dan negatif ini. Loss BPR dihitung berdasarkan selisih skor positif dan negatif, dan gradien dihitung melalui *backpropagation* untuk mengukur seberapa besar setiap parameter model berkontribusi terhadap loss.

3.4.6 Hyperparameter Tuning

Dalam penelitian ini, setelah melakukan pelatihan model *default* yang dijadikan *baseline* penelitian, dilakukan pencarian nilai *hyperparameter* untuk menyesuaikan kombinasi yang optimal bagi model lightGCN dengan menguji berbagai rentang untuk mengukur potensi dari model. Di mana serangkaian nilai potensial untuk *hyperparameter* kunci seperti *embedding dimension*, *learning rate*, jumlah lapisan propagasi, serta *optimizer* diuji secara sistematis dalam semua kombinasinya, pengaturan parameter *embedding dimension* diuji dari 32, 64, 128, 256, dan 512, *learning rate* 0.005, 0.001, 0.0005, 0.0001, 0.00005, dan jumlah lapis propagasi diukur dari rentang 1 hingga 5 lapis, karena penelitian oleh Aldayel dkk., (2023); He dkk., (2020); S. Lee dkk., (2024) masing masing sudah menggunakan 2, 3 dan 4 lapis propagasi, serta *optimizer* yang diuji adalah Adam, RMSProp, SGD, Adagrad dan Nadam (Anggara dkk., 2023)

3.3 Studi Deskriptif 1

Tahap berikutnya, Studi Deskriptif 1 bertujuan untuk merumuskan masalah yang menjadi fokus penelitian serta menentukan tujuan yang ingin dicapai berdasarkan hasil klarifikasi penelitian yang dilakukan sebelumnya, rumusan masalah yang telah ditetapkan menjadi acuan dalam merumuskan *state-of-the-art* penelitian dari penelitian-penelitian terdahulu sebagai tolak ukur celah dari penelitian ini. Setelah merumuskan rumusan masalah selanjutnya dilakukan perumusan tujuan penelitian berdasarkan studi literatur yang sudah dilakukan. Rumusan masalah dan tujuan penelitian ini menjadi dasar dalam mempersiapkan solusi dalam mengatasi tantangan dalam penelitian yang telah dirumuskan.

3.5 Studi Deskriptif 2

Tahap terakhir dalam *Design Research Methodology* adalah Studi Deskriptif 2, yang bertujuan untuk mengevaluasi model yang telah dihasilkan. Evaluasi ini dilakukan guna memperoleh pemahaman mengenai kinerja model yang dibangun. Pengujian dilakukan dengan menggunakan data validasi (10%) guna menentukan konfigurasi parameter terbaik, tanpa melibatkan data uji. Setelah model terbaik didapatkan, evaluasi selanjutnya dilakukan menggunakan data uji (10%), Evaluasi dilakukan untuk mengukur kinerjanya menggunakan metrik-metrik evaluasi yang digunakan dalam sistem rekomendasi khususnya yang berbasis graf, seperti *Recall* dan *Normalized Discounted Cumulative Gain* (NDCG) seperti pada penelitian oleh He dkk. (2020) yang mengevaluasi LightGCN. Hasil dari evaluasi ini dianalisis untuk menentukan sejauh mana model mampu meningkatkan akurasi dan relevansi rekomendasi destinasi wisata pada data uji.

3.6 Alat dan Bahan Penelitian

Penelitian ini menggunakan berbagai alat dan bahan untuk mendukung proses implementasi, pengembangan, dan pengujian sistem rekomendasi destinasi wisata. Alat-alat yang digunakan mencakup perangkat keras, perangkat lunak, Bahasa Pemrograman, Dataset serta *Library* yang relevan dengan kebutuhan pengembangan. Tujuan dari penggunaan alat dan bahan ini adalah untuk

memastikan kelancaran proses pengembangan model. *Tabel 3.* di bawah ini memuat daftar spesifikasi alat dan bahan yang digunakan selama penelitian ini:

Tabel 3. 1 Konfigurasi perangkat keras dan perangkat lunak

Jenis Perangkat	Tipe
<i>Processor</i>	Intel® Core™ i5-1035G1 @ 1.00 GHz
RAM	8 GB
SSD	239 GB
Sistem Operasi	Windows 10
<i>Python</i>	3.10
IDE	Google Colaboratory

Tabel 3. 2 Daftar library

Nama Library	Kegunaan
<i>Pandas</i>	Untuk memuat, manipulasi, pembersihan, dan eksplorasi data dalam struktur DataFrame.
<i>Matplotlib</i>	Untuk membuat visualisasi dasar (histogram, line plot), mengatur ukuran dan tampilan figure.
<i>Seaborn</i>	Membangun di atas matplotlib untuk plot statistik yang lebih cantik.
<i>Scikit-learn</i>	Menyediakan fungsi <code>train_test_split</code> untuk membagi dataset menjadi train/validation/test.
<i>NumPy</i>	Operasi numerik (perhitungan DCG, IDCg, array manipulations) yang dipakai di evaluasi metrik.

<i>PyTorch</i>	Framework deep learning untuk Representasi data sebagai tensor, Membuat dan menginisialisasi embedding, Definisi model LightGCN, Operasi sparse-matrix multiplication untuk message passing.
<i>Itertools</i>	untuk membuat kombinasi hyperparameter grid search
<i>Random</i>	Untuk sampling negatif: memilih user/item secara acak yang belum berinteraksi.

3.7 Metrik Penelitian

Metrik evaluasi yang digunakan dalam Penelitian ini adalah metrik metrik standar yang digunakan dalam menilai kinerja sistem rekomendasi seperti pada penelitian terdahulu (Aldayel dkk., 2023; S. Lee dkk., 2024). seluruh metrik memiliki skala 0 hingga 1, menunjukkan semakin mendekati 1 maka semakin baik performa dan akurasi model, berikut adalah metrik evaluasi yang digunakan:

1. ***Recall***:

$$\frac{Tp}{Tp + F_n}$$

Di mana :

- Tp = Jumlah kasus positif yang di prediksi benar secara positif.
- F_n = Jumlah kasus positif yang salah diprediksi sebagai negative (model gagal mendeteksi nya)

2. ***Normalized Discounted Cumulative Gain (NDCG)***:

$$NDCG = \frac{DCG}{IDCG}$$

Frahari Perdana Putra, 2025

OPTIMASI MODEL LIGHT GRAPH CONVOLUTIONAL NETWORK PADA REKOMENDASI DESTINASI WISATA BERDASARKAN RIWAYAT INTERAKSI PENGGUNA

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

- DCG = Jumlah kumulatif gain (skor relevansi) yang terdiskon berdasarkan posisi item dalam urutan, sehingga item di peringkat atas mendapat bobot lebih tinggi.
- $IDCG$ = Nilai maksimum (ideal), yaitu DCG yang diperoleh apabila semua item diurutkan dari relevansi tertinggi ke terendah.

3.8 Analisis Data

Pada tahap analisis data dilakukan setelah mendapat hasil uji berdasarkan metrik metrik evaluasi yang telah ditetapkan untuk mengidentifikasi bagaimana kinerja model yang telah diimplementasikan. Analisis data ini juga mencakup perbandingan kinerja model LightGCN untuk sistem rekomendasi wisata dengan berbagai rentang lapis propagasi sehingga Setelah menganalisis lapis propagasi model diharapkan mampu mendapatkan hasil terbaik berdasarkan metrik evaluasi yang telah ditetapkan dimana metrik-metrik evaluasi tersebut adalah Recall, dan *Normalized Discounted Cumulative Gain* (NDCG).