BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil perancangan, implementasi, dan evaluasi sistem rekomendasi end-to-end yang mengintegrasikan SASRec, Large Language Model (LLM), dan Inverse Propensity Scoring (IPS), maka kesimpulan penelitian ini dirumuskan dengan menjawab rumusan masalah sebagai berikut:

1. Kinerja Model SASRec dalam Menangkap Pola Urutan Interaksi Pengguna

   Model SASRec terbukti efektif dalam memodelkan dinamika preferensi pengguna berbasis urutan interaksi historis. Hal ini ditunjukkan oleh capaian HR@10 sebesar 84,56%, NDCG@10 sebesar 0,7158, dan AUC sebesar 0,9495 pada test set. Mekanisme self-attention mampu menangkap dependensi jangka pendek maupun jangka panjang dalam sekuens interaksi pengguna secara akurat. Namun demikian, evaluasi fairness menunjukkan adanya bias popularitas yang signifikan, ditandai dengan item coverage sebesar 47,66%, Gini Index sebesar 0,9048, serta long-tail exposure sebesar 19,73%. Temuan ini menunjukkan bahwa meskipun model memiliki akurasi tinggi, distribusi rekomendasi masih terkonsentrasi pada item populer sehingga belum mencerminkan sistem yang adil dan beragam.

   Dengan demikian, rumusan masalah pertama dapat dijawab bahwa SASRec efektif dalam menangkap pola urutan interaksi pengguna dan menghasilkan akurasi tinggi, tetapi masih memiliki permasalahan bias popularitas yang signifikan.

2. Pengaruh Integrasi Large Language Model (LLM) melalui End-to-End Embedding Fusion

   Integrasi representasi semantik berbasis BERT melalui mekanisme gated fusion memberikan peningkatan kualitas rekomendasi dibandingkan baseline. HR@10 meningkat sebesar 0,86% menjadi 85,29%, NDCG@10 meningkat sebesar

0,56% menjadi 0,7198, dan AUC meningkat sebesar 0,17% menjadi 0,9511. Hasil ini menunjukkan bahwa representasi semantik dari informasi tekstual item (judul dan genre) mampu memperkaya pemahaman model di luar pola kolaboratif murni. Mekanisme gated fusion memungkinkan model menentukan kontribusi optimal antara embedding berbasis ID dan embedding semantik secara adaptif. Akan tetapi, dampak integrasi LLM terhadap aspek fairness relatif terbatas. Item coverage hanya meningkat sebesar 4,76%, Gini Index turun sebesar 0,19%, dan long-tail exposure meningkat sebesar 0,20%. Dengan demikian, integrasi LLM efektif meningkatkan akurasi, tetapi belum mampu mengatasi bias struktural dalam distribusi data pelatihan.

Dengan demikian, rumusan masalah kedua dapat dijawab bahwa integrasi LLM melalui pendekatan end-to-end embedding fusion berpengaruh positif terhadap peningkatan akurasi rekomendasi, namun belum secara signifikan mengurangi bias struktural dalam sistem.

3. Penerapan Inverse Propensity Scoring (IPS) untuk Mitigasi Bias

Penerapan IPS dalam fungsi loss terbukti efektif dalam memitigasi bias popularitas dan bias eksposur pada sistem rekomendasi end-to-end. Peningkatan fairness terjadi secara signifikan, dengan item coverage naik menjadi 61,99% (+30,07%), long-tail exposure meningkat menjadi 28,51% (+44,50%), Gini Index menurun menjadi 0,8605 (-4,90%), serta entropy meningkat sebesar 4,36%. Trade-off terhadap akurasi relatif kecil dan masih berada pada tingkat kompetitif, dengan HR@10 turun sebesar 1,54% menjadi 83,98%, NDCG@10 turun sebesar 1,61%, dan AUC turun sebesar 0,45%. Penurunan ini merupakan konsekuensi logis dari proses debiasing karena metrik konvensional dievaluasi pada test set yang juga mengandung bias popularitas.

Dengan demikian, rumusan masalah ketiga dapat dijawab bahwa penerapan Inverse Propensity Scoring (IPS) efektif dalam meningkatkan fairness sistem

rekomendasi dengan trade-off akurasi yang relatif minimal dan masih dalam batas yang dapat diterima.

5.2 Saran

Berdasarkan hasil penelitian dan keterbatasan yang ditemukan, beberapa saran untuk pengembangan penelitian selanjutnya adalah:

1. Eksplorasi Arsitektur LLM yang Lebih Besar

   Mengeksplorasi model seperti BERT-large, RoBERTa, atau GPT untuk representasi semantik yang lebih kaya, serta mempertimbangkan domain-specific language model yang dilatih pada corpus khusus (misalnya data film) untuk meningkatkan kualitas representasi.

2. Pengembangan Mekanisme Fusion yang Lebih Canggih

   Mengeksplorasi mekanisme fusion kompleks seperti multi-head fusion, attention-based fusion, atau cross-attention mechanism untuk interaksi yang lebih kaya antara representasi ID-based dan text-based.

3. Penerapan Metode Debiasing Tambahan

   Mengeksplorasi kombinasi IPS dengan metode lain seperti Doubly Robust estimation, causal embedding, atau adversarial debiasing, serta mengembangkan mekanisme debiasing adaptif yang menyesuaikan tingkat debiasing berdasarkan karakteristik pengguna.

4. Evaluasi pada *Dataset* yang Lebih Beragam

   Mengintegrasikan mekanisme seperti attention visualization, gradient-based explanation, atau natural language explanation untuk meningkatkan transparansi dan kepercayaan pengguna.

5. Integrasi dengan Mekanisme Explainability

   Mengintegrasikan mekanisme seperti attention visualization, gradient-based explanation, atau natural language explanation untuk meningkatkan transparansi dan kepercayaan pengguna.

6. Evaluasi dengan User Study

   Melakukan online evaluation atau user study untuk mengukur dampak terhadap kepuasan pengguna, user experience, perceived fairness, dan long-term engagement melalui A/B testing dalam kondisi real-world.