

ABSTRAK

Pertumbuhan data di era digital menimbulkan masalah kelebihan informasi (*information overload*) bagi wisatawan dalam memilih destinasi kuliner yang sesuai dengan preferensi pribadi. Pada konteks rekomendasi tempat kuliner, pendekatan *Collaborative Filtering* (CF) sering menghadapi kendala kelangkaan data (*sparsity*) dan masalah pengguna baru (*cold-start*) karena bergantung pada riwayat interaksi pengguna. Sementara itu, *Content-Based Filtering* (CBF) memiliki keterbatasan dalam menangkap konteks preferensi yang lebih luas karena rekomendasi didasarkan pada kemiripan fitur item. Selain itu, penggunaan *rating* numerik saja belum mampu merepresentasikan opini pengguna secara menyeluruh. Permasalahan tersebut menunjukkan perlunya pendekatan untuk mengatasi kelemahan metode tunggal seperti *sparsity*, *cold-start*, dan keterbatasan konteks preferensi, diperlukan pendekatan hibrida yang mengintegrasikan CF, CBF, dan Analisis Sentimen.

Penelitian ini menggunakan data tempat kuliner di Daerah Istimewa Yogyakarta yang diperoleh melalui proses *scraping* Google Maps dengan total 1.428.664 ulasan mentah. Pendekatan yang digunakan adalah *Hybrid Filtering* yang mengombinasikan *Content-Based Filtering*, *Collaborative Filtering*, dan Analisis Sentimen. Pada pendekatan CBF, representasi fitur tempat kuliner diperoleh menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan dihitung kemiripannya menggunakan *Cosine Similarity*. Pada pendekatan CF, prediksi rekomendasi dilakukan menggunakan algoritma *Extended K-Nearest Neighbors Collaborative Filtering* (ExtKNNCF) berbasis parameter kognisi pengguna. Analisis sentimen dilakukan menggunakan model IndoBERT untuk memperoleh skor sentimen dari ulasan pengguna. Skor rekomendasi kemudian digabungkan melalui mekanisme *Switching Hybrid* untuk mengombinasikan CF dan CBF, serta *Weighted Hybridization* untuk mengintegrasikan hasil hibrida dasar dengan skor sentimen. Evaluasi kinerja model dilakukan menggunakan metrik *Hit Rate* (HR@K) dan *Mean Reciprocal Rank* (MRR@K).

Hasil penelitian menunjukkan bahwa *dataset* kuliner memiliki tingkat *sparsity* ekstrem sebesar 0,996246 dengan dominasi pengguna *cold-start* sebesar 80,9%. Kondisi tersebut menyebabkan model CF murni memperoleh kinerja terendah dengan nilai rata-rata HR sebesar 0,0626 dan MRR sebesar 0,0155. Model CBF menunjukkan kinerja yang lebih stabil dengan nilai rata-rata HR sebesar 0,4726 dan MRR sebesar 0,2119. Performa terbaik diperoleh model *Weighted Hybrid* dengan nilai rata-rata keseluruhan HR sebesar 0,4853 dan MRR sebesar 0,2342. Capaian tersebut diperoleh pada kombinasi bobot 0,2 untuk komponen sentimen dan 0,8 untuk komponen hibrida dasar. Berdasarkan hasil evaluasi tersebut, integrasi Analisis Sentimen pada pendekatan *Hybrid Filtering* mampu meningkatkan akurasi dan kualitas pemeringkatan rekomendasi kuliner, terutama pada kondisi data yang memiliki *sparsity* tinggi, dominasi pengguna *cold-start*, dan keterbatasan konteks preferensi pengguna.

Kata Kunci: Sistem Rekomendasi, *Hybrid Filtering*, *Content-Based Filtering*, *Collaborative Filtering*, Analisis Sentimen, IndoBERT, *Cold-Start*.

ABSTRACT

The rapid growth of data in the digital era has created an information overload problem for tourists when selecting culinary destinations that match their personal preferences. In the context of culinary place recommendation, Collaborative Filtering (CF) often faces data sparsity and cold-start problems because it relies heavily on users' historical interactions. Meanwhile, Content-Based Filtering (CBF) has limitations in capturing broader preference contexts because recommendations are generated based on item feature similarity. In addition, the use of numerical ratings alone is not sufficient to fully represent users' opinions, particularly when inconsistencies occur between rating values and textual reviews. These problems indicate the need for a recommendation approach that can integrate user interaction patterns, culinary place characteristics, and sentiment information from reviews to improve recommendation accuracy and relevance.

This research uses culinary place data in the Special Region of Yogyakarta obtained through Google Maps scraping, resulting in 1,428,664 raw reviews. The approach used in this research is Hybrid Filtering, which combines Content-Based Filtering, Collaborative Filtering, and Sentiment Analysis. In the CBF approach, culinary place feature representation is generated using Term Frequency-Inverse Document Frequency (TF-IDF), and similarity is calculated using Cosine Similarity. In the CF approach, recommendation prediction is performed using the Extended K-Nearest Neighbors Collaborative Filtering (ExtKNNCF) algorithm based on user cognition parameters. Sentiment analysis is performed using the IndoBERT model to obtain sentiment scores from user reviews. The recommendation scores are then combined through a Switching Hybrid mechanism to integrate CF and CBF, and Weighted Hybridization is used to integrate the base hybrid results with sentiment scores. Model performance is evaluated using Hit Rate (HR@K) and Mean Reciprocal Rank (MRR@K).

The results show that the culinary dataset has an extreme sparsity level of 0.996246, with cold-start users dominating at 80.9%. This condition causes the pure CF model to obtain the lowest performance, with an average HR of 0.0626 and an average MRR of 0.0155. The CBF model demonstrates more stable performance, with an average HR of 0.4726 and an average MRR of 0.2119. The best performance is achieved by the Weighted Hybrid model, with an overall average HR of 0.4853 and an average MRR of 0.2342. This result is obtained using a weight combination of 0.2 for the sentiment component and 0.8 for the base hybrid component. Based on these evaluation results, the integration of Sentiment Analysis into the Hybrid Filtering approach is able to improve the accuracy and ranking quality of culinary recommendations, especially under conditions of high data sparsity, dominant cold-start users, and limited user preference context.

Keywords: *Recommender System, Hybrid Filtering, Content-Based Filtering, Collaborative Filtering, Sentiment Analysis, IndoBERT, Cold-Start.*