

ABSTRAK

Meningkatnya jumlah buku yang beredar akan menyulitkan dalam menemukan sebuah buku yang relevan sesuai dengan minat dan kebutuhan seseorang. Penggunaan sistem rekomendasi dapat menjadi solusi permasalahan tersebut dengan menghasilkan rekomendasi sesuai dengan preferensi atau kesukaan penggunanya. Ada beberapa metode sistem rekomendasi yang dikenal, yaitu *content based filtering*, *collaborative filtering* dan *hybrid*. *Content based filtering* dapat memberikan rekomendasi item tanpa adanya parameter *rating*, namun *content based filtering* memiliki masalah utama yaitu dapat menghasilkan item rekomendasi yang kurang beragam. Penggunaan *collaborative filtering* dapat memberikan rekomendasi yang lebih beragam, tetapi metode ini masih memiliki kekurangan berupa *cold start* yaitu tidak dapat menghasilkan rekomendasi dikarenakan tidak memiliki informasi yang cukup tentang data *rating* dari suatu item. Pembuatan sistem rekomendasi memerlukan model rekomendasi yang tepat karena selain akurasi, keberagaman rekomendasi juga menentukan kualitas dari sistem rekomendasi. Penggabungan metode *content based filtering* dan *collaborative filtering* menjadi *hybrid* akan memberikan hasil rekomendasi yang baik dengan memberikan daftar rekomendasi yang lebih beragam namun tetap sesuai dengan minat dan kebutuhan dari *user*. Hal ini dikarenakan metode-metode yang digabungkan akan saling mengatasi kekurangan dari metode satu sama lain.

Data yang digunakan penelitian ini adalah data “GoodBook10k” yang berisi 2 jenis data yaitu data buku dan data *rating*. Model yang dibangun adalah model *hybrid* dengan penggabungan dari model *content based filtering* dan *collaborative filtering* secara linear. Sebelum model dibangun, data buku akan dilakukan proses *preprocessing* data berupa modifikasi tabel, *case folding*, *cleansing*, *tokenization*, *stopwords removal*, dan *stemming*. Data buku hasil *preprocessing* data kemudian digunakan untuk pembuatan model *content based filtering* dengan cara menghitung bobot pada setiap *term* pada buku dengan *TF-IDF* dan menghitung similaritas tiap buku dengan *cosine similarity*, sehingga nilai similaritas tertinggi digunakan sebagai acuan dalam memberikan rekomendasi. Kemudian untuk data *rating* digunakan untuk pembuatan model *collaborative filtering* dengan cara melakukan *split* data dengan perbandingan 80% data latih dan 20% data uji lalu melakukan *training* model menggunakan algoritma *cosine similarity* dan *k-Nearest Neighbors (KNN)* untuk mendapatkan prediksi *rating* dari item sebagai acuan dalam memberikan rekomendasi. Lalu hasil rekomendasi dari kedua model digabung menjadi model *hybrid* untuk dapat menghasilkan rekomendasi akhir yang diterima pengguna.

Pengujian dilakukan dengan menggunakan *Mean Absolute Error (MAE)* untuk mengukur tingkat *error* pada model dalam memprediksi *rating*, semakin kecil nilai yang dihasilkan menunjukkan model yang semakin baik dan menghasilkan parameter *k* paling optimal yang digunakan pada model. Kemudian pengujian dengan *Intra-list Similarity* untuk mengukur keberagaman yang dihasilkan oleh model, semakin kecil nilai yang dihasilkan maka rekomendasi yang dihasilkan oleh model semakin beragam. Hasil penelitian menunjukkan bahwa nilai *MAE* terbaik pada $k = 40$ yaitu sebesar 0,682. Kemudian nilai *Intra-list Similarity* terbaik ada pada top-10 rekomendasi model *hybrid* yaitu sebesar 0,532.

Kata Kunci : Buku, Sistem Rekomendasi, Hybrid

ABSTRACT

The increasing number of books in circulation will make it difficult to find a book that is relevant to one's interests and needs. The use of a recommendation system can be a solution to this problem by producing recommendations according to the preferences of the users. There are several known recommendation system methods, namely content based filtering, Collaborative Filtering and hybrid. Content-based filtering can provide item recommendations without rating parameters, but content-based filtering has the main problem, namely it can produce item recommendations that are less diverse. The use of Collaborative Filtering can provide a more diverse recommendation, but this method still has the disadvantage of cold start, namely it cannot produce recommendations because it does not have sufficient information about the rating data of an item. Making a recommendation system requires an appropriate recommendation model because apart from accuracy, the diversity of recommendations also determines the quality of the recommendation system. Combining the content-based filtering and Collaborative Filtering methods into a hybrid will provide good recommendation results by providing a more diverse list of recommendations but still according to the interests and needs of the users. This is because the combined methods will mutually overcome the shortcomings of each other's methods.

The data used in this study is "GoodBook10k" data which contains 2 types of data, namely book data and rating data. The model built is a hybrid model with a linear merging of content-based filtering and Collaborative Filtering models. Before the model is built, the book data will be preprocessed by data preprocessing in the form of table modification, case folding, cleaning, tokenization, removing stopwords, and stemming. Book data resulting from data preprocessing is then used to create a content-based filtering model by calculating the weight of each term in the book with TF-IDF and calculating the similarity of each book with cosine sameness, so that the highest similarity value is used as a reference in providing recommendations. Then the rating data is used to create a collaborative filtering model by splitting the data with a ratio of 80% training data and 20% test data and then carrying out a training model using the cosine similarity algorithm and k-Nearest Neighbors (KNN) to get the rating predictions from the item as a reference. in providing recommendations. Then the recommendations from the two models are combined into a hybrid model to produce final recommendations that are accepted by the user.

Testing is carried out using the Mean Absolute Error (MAE) to measure the error rate in the model in predicting ratings, the smaller the resulting value indicates the better the model and produces the most optimal k parameters used in the model. Then testing with the Intra-list of Similarities to measure the diversity produced by the model, the smaller the value generated, the more diverse the recommendations produced by the model. The results showed that the best MAE value was at $k = 40$ which was 0.682. Then the best Intra-list Similarity value is in the top-10 recommendations for the hybrid model, namely 0.532.

Keywords: Books, Recommendation System, Hybrid