

BAB I PENDAHULUAN

1.1 Latar Belakang

Sistem rekomendasi merupakan sistem yang dibangun untuk penyelesaian suatu prediksi *item* sesuai dengan minat dari pengguna [1]. Dalam sistem rekomendasi didasarkan pada pemberian peringkat untuk mewakili preferensi yang dilakukan oleh pengguna terhadap beberapa *item* atau pemberian peringkat dari beberapa pengguna terhadap suatu *item* [2]. Pemberian peringkat dalam sistem rekomendasi biasanya hanya mempertimbangkan satu kriteria nilai, tetapi terdapat beberapa data yang memiliki lebih dari satu kriteria (multi-kriteria). Dengan adanya multi-kriteria dapat membantu meningkatkan kualitas rekomendasi, karena multi-kriteria dapat mewakili preferensi pengguna dengan lebih baik dibandingkan dengan kriteria tunggal [3].

Multi-kriteria memiliki tingkat keakuratan yang lebih tinggi dibandingkan dengan kriteria tunggal karena preferensinya lebih lengkap. Namun multi-kriteria juga memiliki beberapa permasalahan, yang pertama cara penyelesaian untuk mendapatkan rekomendasi dalam multi-kriteria berbeda dengan kriteria tunggal [4]. Yang kedua, kriteria utama dan subkriteria dalam sistem rekomendasi multi-kriteria mempunyai skala penilaian yang berbeda sehingga memperumit permasalahan karena pemahaman pengguna tentang skala penilaian antar kriteria berbeda [4].

Dari permasalahan multi-kriteria tersebut, terdapat beberapa cara penyelesaian yang telah digunakan oleh para peneliti dalam melakukan rekomendasi *item* terhadap pengguna, diantaranya pendekatan *Collaborative Filtering* (CF) yang terdiri dari *Memory-based*(*user-based*, *item-based*) dan *Model-based* [5]. Metode CF yang sering digunakan dalam penelitian yaitu CF berbasis model [6], CF dengan *Content-based* [1], serta *user-based* CF (UBCF) dikombinasikan dengan *item-based* CF (IBCF) [7]. Metode kombinasi UBCF dan IBCF dinilai lebih bagus dibandingkan dengan metode lainnya, karena penggabungan UBCF dan IBCF menghasilkan nilai akurasi yang lebih baik [7], serta proses perhitungan UBCF dan IBCF lebih mudah dibandingkan dengan metode lainnya.

Selain permasalahan multi-kriteria terdapat juga permasalahan lain yaitu tentang skala penilaian. Dimana skala penilaian yang berbeda antara pengguna

satu dengan yang lainnya menyebabkan nilai *rating* tidak seimbang dikarenakan ada nilai yang terlalu tinggi dan ada nilai yang terlalu rendah [4]. Ketidakseimbangan *rating* tersebut akan mempengaruhi hasil rekomendasi yang akan diberikan terhadap pengguna karena rekomendasi yang didapatkan oleh pengguna kurang sesuai bahkan tidak sesuai dengan kebutuhannya, sehingga pengguna tidak puas dengan hasil dari rekomendasi yang diberikan oleh sistem. Oleh karena itu, dibutuhkan metode untuk mengatasi permasalahan skala penilaian tersebut.

Skala penilaian dapat diatasi dengan metode normalisasi *rating*, dimana normalisasi *rating* digunakan untuk mengatasi nilai *rating* yang terlalu tinggi atau terlalu rendah, serta menghasilkan struktur data yang normal. Sehingga dengan adanya normalisasi *rating*, data atau nilai *rating* yang akan digunakan untuk perhitungan rekomendasi menjadi lebih stabil yang nantinya akan mempengaruhi tingkat akurasi dari sistem [4]. Ada peneliti yang menggunakan metode normalisasi *rating* (*z-score* dan *decoupling*) [4], normalisasi *rating* (*min-max*, *z-score*, *decimal scale*) [8, 9], serta normalisasi *rating* (*decimal scale*, *sigmoid*, *softmax*, *min-max*, *statistical column*, *z-score*) [10]. Dari berbagai metode normalisasi *rating* tersebut, *min-max* merupakan metode dengan nilai akurasi paling tinggi [8] dikarenakan normalisasi *rating min-max* mengambil rentang nilai maksimal dan minimal dari data, sehingga dapat mewakili keseluruhan data yang ada [10].

Pada penelitian ini menggunakan metode Normalisasi *rating* [4] *min-max* [8] serta kombinasi antara UBCF dan IBCF [7]. Metode ini digunakan karena dapat memperbaiki rekomendasi dengan cara normalisasi *rating* dan menghitung prediksi berdasarkan UBCF dan IBCF yang kemudian hasil prediksi UBCF dan IBCF akan dilakukan penggabungan, sehingga dapat mengatasi permasalahan skala penilaian yang dapat merubah data menjadi lebih baik, serta mampu menghasilkan tingkat akurasi yang lebih tinggi. Hasil dari penelitian ini mampu menghasilkan akurasi rekomendasi yang lebih tinggi jika dibandingkan dengan metode normalisasi *rating decoupling* yang dikombinasikan dengan *User-based*.

1.2 Perumusan Masalah

Ada 3 poin penting di dalam perumusan masalah, yaitu masalah penelitian, metode usulan, dan pertanyaan penelitian.

1.2.1 Masalah Penelitian

Penyelesaian multi-kriteria terbilang cukup sulit dikarenakan terdapat beberapa kriteria yang harus diselesaikan. Disisi lain terdapat permasalahan skala penilaian, dimana skala penilaian yang berbeda antara pengguna satu dengan yang lainnya menyebabkan nilai *rating* tidak seimbang. Ketidakseimbangan nilai *rating* mengakibatkan rendahnya nilai akurasi sistem rekomendasi.

1.2.2 Metode Penelitian

Penelitian ini menggunakan metode normalisasi *rating min-max* serta kombinasi antara UBCF dan IBCF untuk menyeimbangkan nilai *rating* dan meningkatkan nilai akurasi sistem rekomendasi.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan dan metode yang telah dijelaskan diatas, didapatkan dua pertanyaan yaitu:

1. Bagaimana akurasi rekomendasi yang dihasilkan oleh metode yang menerapkan teknik normalisasi *rating min-max* dan kombinasi model UBCF & IBCF?
2. Apakah metode yang menerapkan teknik normalisasi *rating min-max* dan kombinasi model UBCF & IBCF lebih bagus daripada metode yang mengkombinasikan teknik normalisasi *decoupling* dan model *User-Based*?

1.3 Tujuan & Manfaat

1.3.1 Tujuan Penelitian

Mengacu pada latar belakang serta rumusan masalah diatas maka tujuan dari penelitian ini adalah:

1. Untuk mengetahui akurasi rekomendasi yang dihasilkan oleh metode yang menerapkan teknik normalisasi *rating min-max* dan kombinasi model UBCF & IBCF.
2. Untuk mengetahui metode yang menerapkan teknik normalisasi *rating min-max* dan kombinasi model UBCF & IBCF lebih bagus daripada metode yang mengkombinasikan teknik normalisasi *decoupling* dan model *User-Based*.

1.3.2 Manfaat Penelitian

Hasil dari penelitian ini mampu menghasilkan akurasi rekomendasi yang lebih tinggi jika dibandingkan dengan metode normalisasi *rating decoupling* yang dikombinasikan dengan *User-based*.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Dataset yang digunakan dalam penelitian ini berupa data multi kriteria yang didapat dari Yelp Dataset (<https://www.kaggle.com/yelp-dataset/yelp-dataset/version/6>).
2. Yelp Dataset yang digunakan berupa 25.000 data *rating* dengan empat kriteria untuk hotel yaitu (*Overall, Usefull, Funny, Cool*).
3. Sistem rekomendasi dalam penelitian ini tidak untuk *user* baru atau tidak menghitung *cold-start problem* baik pada *user* maupun *item*.

1.5 Metode Penelitian

Secara umum, penelitian dilakukan dengan beberapa tahapan, yaitu:

a. Studi Pustaka

Pada tahap ini, semua informasi yang diperlukan dalam penelitian akan dikumpulkan dan dipelajari melalui media internet, buku, maupun jurnal. Data atau informasi yang dikumpulkan berupa materi tentang sistem rekomendasi multi-kriteria, normalisasi *rating*, kombinasi *user-based* dan *item-based collaborative filtering*, dan penelitian terkait. Selain itu, dataset yang akan digunakan pada penelitian ini adalah Yelp Dataset.

b. Analisa dan Perancangan Sistem

Analisa permasalahan diperoleh setelah mempelajari semua data yang telah dikumpulkan. Dari data-data tersebut didapatkan permasalahan bahwa sistem rekomendasi multi-kriteria mempunyai skala penilaian yang berbeda antar kriteria, sehingga menyebabkan nilai *rating* yang diberikan antara pengguna satu dengan pengguna lain tidak seimbang. Ketidakseimbangan *rating* berpengaruh pada nilai akurasi dari sistem, oleh karena itu beberapa peneliti telah melakukan riset agar permasalahan skala penilaian dapat diselesaikan dan dapat menghasilkan nilai akurasi yang tinggi.

c. Implementasi Sistem

Tahap implementasi dilakukan setelah membuat rancangan sistem. Implementasi dimulai dengan mengubah dataset menjadi matrik *rating*, kemudian membuat program normalisasi *rating min-max*. Hasil dari normalisasi *rating min-max* akan digunakan pada program untuk menghitung prediksi berdasarkan UBCF dan IBCF. Kemudian hasil prediksi UBCF dan IBCF akan digunakan pada program penggabungan UBCF dan IBCF dengan teknik *Fusion*. Dari proses penggabungan tersebut akan menghasilkan prediksi *rating* yang kemudian akan mendapatkan Top-N rekomendasi untuk diberikan kepada pengguna.

d. Uji Coba Sistem

Pada tahap ini, dilakukan proses uji coba menggunakan data pelatihan, dimana data pelatihan tersebut akan diterapkan pada metode normalisasi *rating min-max* yang dikombinasikan dengan UBCF dan IBCF, sehingga akan menghasilkan prediksi *rating* yang kemudian akan mendapatkan Top-N rekomendasi untuk diberikan kepada pengguna.

e. Evaluasi

Top-N rekomendasi yang dihasilkan dari proses uji coba, selanjutnya akan dilakukan proses evaluasi menggunakan *F1-Score* untuk mengetahui nilai akurasi dari metode normalisasi *rating min-max* yang dikombinasikan dengan UBCF dan IBCF. Pada tahap evaluasi akan diketahui seberapa akurat metode yang digunakan pada penelitian ini.

f. Dokumentasi

Tahap dokumentasi dilakukan mulai dari tahap studi pustaka hingga tahap evaluasi.

1.6 Sistematika Penulisan

Sistematika penulisan Proposal ini terdiri dari beberapa bab, yaitu :

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, metode penelitian, serta sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang landasan teori dan penelitian sebelumnya.

BAB III METODE USULAN

Pada bab ini membahas tentang metode yang diusulkan (algoritma), kebutuhan *software* dan *hardware* yang dibutuhkan, arsitektur (Input, Proses, Ouput), skenario pengujian, serta dataset.

BAB IV IMPLEMENTASI METODE

Pada bab ini menjelaskan implementasi dari metode usulan yang akan dibuat pada bab sebelumnya, dimana sistem akan diuji coba berdasarkan model dan skenario yang telah ditentukan.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dari hasil uji coba sistem dan saran untuk pengembangan penelitian terkait selanjutnya.