**Abstrak**

Item-based collaborative filtering adalah teknik yang lebih disukai pada sistem rekomendasi. Ini menggunakan nilai kesamaan peringkat item untuk memprediksi preferensi pengguna. Dalam makalah ini, kami menyertakan nilai-nilai kesamaan atribut item untuk menyesuaikan persamaan peringkat prediksi untuk item target. Hasil pemfilteran kolaboratif berbasis Item bahwa teknik kesamaan item item dan kemiripan atribut atribut item memiliki Mean Absolute Error (MAE) kurang dari teknik penyaringan kolaboratif berbasis Item tradisional dan lain-lain. Algoritma yang diusulkan efisien untuk memprediksi lebih baik daripada algoritma tradisional seperti yang ditunjukkan dalam percobaan kami

**Perkenalan**

Saat ini, Internet adalah hal penting dalam kehidupan dan mengumpulkan informasi untuk pengguna sistem, dengan penyebab agregasi komunitas yang sangat penting. E-commerce adalah bisnis yang menawarkan produk mereka kepada pengguna untuk melakukan pembelian seperti makanan, film, dan buku, dll. Ada banyak metode yang berbeda untuk membantu rujukan. Sistem rekomendasi [2] adalah salah satu teknologi yang telah membantu memperkenalkan produk untuk memenuhi kebutuhan pelanggan. Ada banyak sistem E-niaga seperti Amazon, CDnow, Netflix, MovieLens, dll yang menggunakan algoritme rekomendasi yang berbeda, tetapi semua sistem akan menawarkan produk yang disarankan kepada pengguna ketika pengguna mengakses sistem. Ada dua teknik, penyaringan berbasis konten dan penyaringan kolaboratif yang umum digunakan dalam sistem rekomendasi. Metode penyaringan berbasis konten akan menganalisis konten dari setiap item yang digunakan pengguna untuk memberikan peringkat. Sistem akan merekomendasikan item yang memiliki konten yang cocok dengan profil pengguna. Masalah teknik ini adalah hasil rekomendasi akan tergantung hanya pada profil pengguna. Karenanya item yang tidak cocok dengan profil pengguna jarang dipilih untuk direkomendasikan kepada pengguna walaupun item tersebut mungkin menarik. Untuk mengatasi masalah ini, teknik penyaringan kolaboratif dikembangkan dan digunakan secara luas sekarang. Penyaringan kolaboratif [3] pada awalnya dilaksanakan oleh Xerox Parc Institute untuk penelitian, yang disebut "Tapestry". Ada dua teknik, pemfilteran kolaboratif berbasis pengguna [4] dan pemfilteran kolaboratif berbasis item. Penyaringan kolaboratif berbasis pengguna [5] telah membawa peringkat item yang disediakan oleh pengguna untuk dibandingkan dengan peringkat pengguna lain pada item yang sama. Sistem akan membuat prediksi otomatis tentang minat pengguna dengan mengumpulkan preferensi atau merasakan informasi dari banyak pengguna dan menghitung kesamaan antara pengguna.

Namun, teknik ini masih memiliki beberapa masalah tentang jumlah informasi dalam sistem sehingga membutuhkan waktu lama untuk diproses, yang disebut masalah skalabilitas. Kemudian pemfilteran kolaboratif Itembased diusulkan untuk memecahkan masalah pemfilteran kolaboratif berbasis pengguna. Teknik ini menyerupai langkah proses pemfilteran kolaboratif berbasis pengguna tetapi peringkat item ditentukan oleh pengguna yang dulu menyediakannya di masa lalu. Dan tekniknya menggunakan kesamaan antara item untuk memilih item tetangga untuk memprediksi preferensi pengguna pada item target. Hasil pemfilteran kolaboratif berbasis Item lebih efektif dalam memprediksi daripada pemfilteran kolaboratif berbasis pengguna dan juga biasanya diterapkan untuk sistem rekomendasi. Namun, teknik ini masih memiliki masalah karena peringkat item pada sistem tidak tercakup, yang disebut masalah sparsity. Masalah lain adalah item baru yang belum menjadi skor penilaian, yang disebut masalah first-rater. Ketika peringkat jarang atau kurang, itu akan berpengaruh untuk menemukan item tetangga dan kinerja peringkat prediksi menurun.

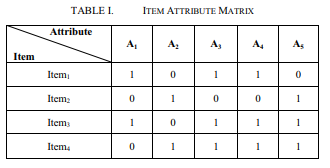
Dalam makalah ini, kami mengusulkan algoritma inovasi untuk meningkatkan efisiensi untuk teknik penyaringan kolaboratif berbasis item. Metode kami akan menggunakan properti item untuk menghitung nilai yang sama. Kami menyebut kesamaan item-atribut, dan kemudian kami menggunakannya untuk menggabungkan dengan nilai kesamaan item dan memprediksi item yang disarankan kepada pengguna target. Hasil dalam prediksi kami lebih akurat setelah kami menemukan kelompok item tetangga dengan memeriksa atribut kesamaan dan menambahkannya ke dalam persamaan prediksi. Oleh karena itu menunjukkan bahwa kesamaan atribut item, atribut item dapat menunjukkan atau menentukan informasi item, dapat digunakan untuk meningkatkan efisiensi seperti yang ditunjukkan dalam percobaan kami. Hasilnya dapat menunjukkan bahwa algoritma yang kami usulkan untuk sistem rekomendasi lebih efektif.

**Pekerjaan yang Terkait dan Latar Belakang**

Teknik penyaringan kolaboratif berbasis Item memiliki berbagai algoritma dan banyak peneliti yang tertarik untuk membuatnya lebih berkualitas. Pengelompokan item diteliti oleh YiBo Huang [6]. Mereka meningkatkan pemfilteran kolaboratif berbasis item dengan menggunakan algoritma pengelompokan k-mean [7] untuk membuat pengelompokan item untuk prediksi dan menggunakan item pengelompokan untuk menemukan kesamaan item dan mampu memprediksi untuk item target.

HengSong Tan [8] mengklasifikasikan item menggunakan konten atribut item untuk menemukan kesamaan item pada setiap item dan mengategorikan atau mengklasifikasikan item ke peringkat prediksi. Ye Zhang [9] menggunakan genre dataset film yang diimpor dari dataset MovieLens. Mereka mengusulkan bahwa genre film dapat membantu menemukan kesamaan film pada setiap film. Mereka menerapkan metode untuk memprediksi peringkat dengan pemfilteran kolaboratif berbasis pengguna dan pemfilteran kolaboratif berbasis item secara terpisah. Setelah itu mereka membandingkan hasil kesamaan setiap film pada item target dan memilih hasil yang baik untuk disajikan kepada pengguna. Hasilnya meningkatkan akurasi tetapi masih menemukan masalah sparsity. SongJie Gong dan HongWu Ye [10] menemukan item tetangga yang menggunakan kesamaan peringkat item dan kesamaan atribut item. Mereka menggunakan kombinasi linier untuk menggabungkan kedua kesamaan untuk peringkat prediksi ke target pengguna. Kemudian mereka harus memilih nilai bobot yang dioptimalkan antara kesamaan peringkat item dan kesamaan atribut-item. Namun, dalam aplikasi dunia nyata, sangat sulit untuk mendapatkan nilai bobot yang dioptimalkan.

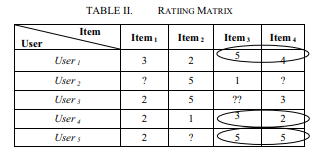
Oleh karena itu, kami mengusulkan algoritma yang hibrid antara kesamaan peringkat item dan kesamaan atribut item dalam persamaan prediksi.



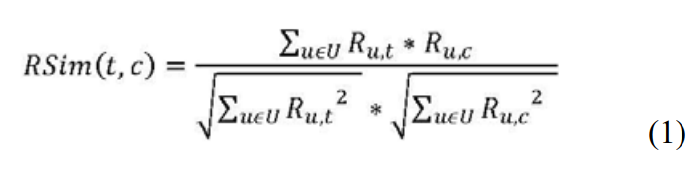
Dalam algoritme yang kami usulkan, kami membuat tabel hubungan antara atribut item sebagai TABEL I. Kami mengklasifikasikan setiap item yang memiliki atribut konten yang sama. Misalnya, item1 memiliki nilai ‘1’ dalam kolom A1, A3 dan A4. Ini berarti item1 terkait dengan atribut A1, A3 dan A4. Dan jika item1 yang tidak terkait dengan atribut A2 dan A5, maka kami mengisi nilai ‘0’ dalam kolom atribut A2 dan A5. Setelah kami mengklasifikasikan semua item, kami akan mendapatkan hasilnya sebagai TABEL I. Kami akan menggunakan semua data dalam TABEL I untuk menghitung kesamaan setiap item menggunakan properti item atribut untuk memprediksi peringkat preferensi untuk pengguna target.

**Metode dan Komputasi**

1. **Item Rating Similarity Method**



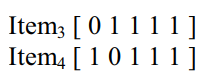
Kami menghitung kesamaan peringkat item dengan membandingkan peringkat item target dengan peringkat semua item. Kami menemukan sekelompok item yang memiliki peringkat yang sama atau yang paling mirip berdasarkan peringkat preferensi. Kami menyebut grup item sebagai nilai bersama seperti yang ditunjukkan pada TABEL II. Metode ini hanya akan mengambil peringkat pemeringkatan untuk menghitung untuk menemukan kesamaan yang cocok antara dua item. Dalam tulisan ini, kami menggunakan metode korelasi kosinus [4] untuk menemukan kesamaan item sebagai (1).



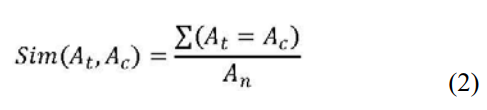
Di mana RSim (t, c) menunjukkan kesamaan antara item t dan c ; Ru,t , Ru,c berarti peringkat pengguna u pada item t dan c masing-masing dan U adalah himpunan pengguna yang memberi peringkat pada kedua item t dan c. Lingkaran di TABEL II menunjukkan peringkat bersama.

1. **Item Attributes Similarity Method**

Metode ini digunakan untuk menemukan kesamaan item dengan menggunakan item atribut. Pada awalnya, kami membawa informasi semua item dalam sistem untuk dibandingkan dengan setiap item. Item yang memiliki kategori atau atribut yang sama akan digunakan untuk membuat matriks atribut Item. Dalam algoritma yang kami usulkan, kami menyesuaikan teknik jarak Hamming [11] untuk diterapkan untuk membandingkan setiap item dan menemukan kesamaan dengan atribut item. Kami memberikan contoh item3 dan item4 untuk dibandingkan untuk menemukan yang serupa sebagai berikut.

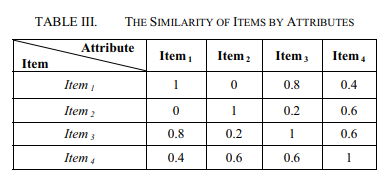


Biasanya metode Hamming Distance menghitung dari sejumlah atribut item yang berbeda yang dibandingkan di masing-masing posisi masing-masing antara Item3 dan Item4 maka hasilnya sama dengan 2. Tapi kami menggunakan sejumlah atribut item yang sama sebagai gantinya. Karena kami ingin menunjukkan kesamaan item dan menunjukkan bahwa dua item erat dan memiliki atribut item yang sama. Kami mengusulkan (2) sebagai.



Di mana Sim (At, Ac) menunjukkan kesamaan atribut antara item t dan c. Pada, Ac berarti nilai atribut A pada item t dan c masing-masing, dan An adalah jumlah semua atribut.

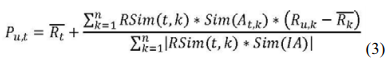
Dalam contoh ini, ketika kita menghitung pasangan atribut item yang sama yang hasilnya sama dengan 3, dan kita melakukan normalisasi dengan membagi dengan jumlah total semua atribut. Kemudian kita bisa membuat tabel hubungan semua item sebagai berikut.



Dalam TABEL III, kami menunjukkan bahwa hasilnya sama dengan 1 jika item dibandingkan dengan dirinya sendiri. Dan jika kita dibandingkan dengan item lain, atribut item tidak tepat. Maka kesamaan item kurang dari 1. Dalam metode yang kami usulkan, kami mendefinisikan Sim (At, k) yang berarti kesamaan atribut item antara item t dan item lainnya dalam sistem.

1. **Prediction by using Item-based Hybrid Similarity**

Pada penelitian kami, kami menyesuaikan metode tradisional untuk memprediksi peringkat sebagai persamaan:



Di mana Pu, t berarti prediksi untuk pengguna u pada item t berdasarkan pada penyaringan kolaboratif berbasis item; RSim (t, k) berarti kesamaan item yang dinilai bersama antara item t dan k; Sim (At, k) menunjukkan kesamaan atribut-item antara item t dan k; Ru,k adalah peringkat pengguna pada item k; Rt Rk adalah peringkat rata-rata pada item t dan k masing-masing, n adalah total tetangga item t; Dari (3), kami menggunakan item kesamaan RSim (t, k) dan kesamaan item-atribut Sim (At, k) untuk meningkatkan akurasi prediksi. Hasilnya adalah peringkat preferensi pengguna yang menyediakan item target dan dapat merekomendasikan item kepada pengguna.

**Eksperimental dan Hasil**

Kumpulan data yang digunakan dalam percobaan adalah dari situs web MovieLens. Situs ini menyimpan data untuk digunakan untuk merekomendasikan film dengan peringkat film pilihan oleh pengguna anggota. Kumpulan data MovieLens dikumpulkan oleh Proyek Penelitian GroupLens di The University of Minnesota. Dataset historis terdiri dari 100.000 peringkat dari 943 pengguna di 1682 film dengan setiap pengguna memiliki setidaknya 20 peringkat. Peringkat mengikuti 1 hingga 5 skala numerik. Jumlah yang berharga berarti pengguna paling suka film itu. Ada 19 genre yang tidak diketahui | Aksi | Petualangan | Animasi | Anak-Anak | Komedi | Kejahatan | Dokumenter | Drama | Fantasi | Film-Noir | Horor | Musik | Misteri | Romantisme | Sains | Sains | Thriller | Perang | Barat. Kemudian kami menggunakan informasi ini untuk membuat matriks atribut Item. Setiap film memiliki 19 genre kemudian sebuah matriks memiliki 19 kolom. Setelah itu kita menghitung kesamaan item-atribut. .

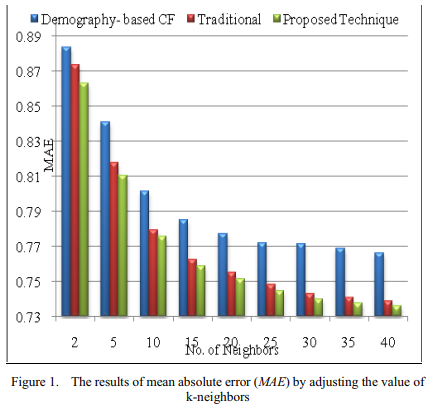
Kami mengevaluasi algoritme kami dengan membandingkan keakuratan prediksi. Kami menggunakan Mean Absolute Error (MAE) [13]; MAE mengembalikan hasil kesalahan dari peringkat prediksi antara peringkat nyata atau peringkat menyembunyikan (data pengujian) dan peringkat prediksi baru. MAE direpresentasikan sebagai berikut.

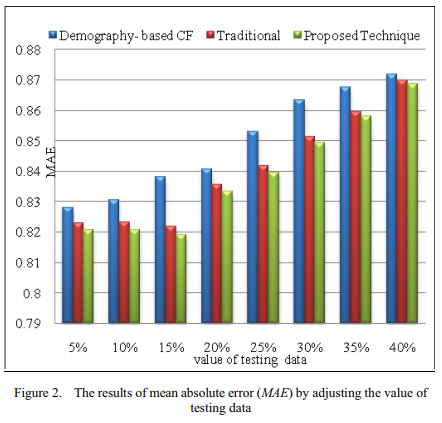


Di mana, pi adalah peringkat yang diprediksi, qi adalah peringkat aktual dan n adalah jumlah peringkat aktual dalam data pengujian. Semakin rendah MAE, semakin akurat prediksi yang memungkinkan untuk memberikan rekomendasi yang lebih baik.

1. **Experiment I**

Kami menguji kinerja dengan menyesuaikan nilai k untuk memiliki jumlah item yang bertetangga 2, 5, 10, 15, 20, 25, 30, 35, dan 40 masing-masing. Kemudian kami memilih film yang memiliki kesamaan dengan film target. Hasilnya ditunjukkan pada Gambar. 1.





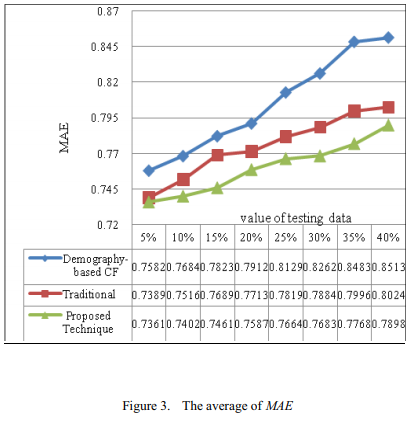
Pada Gambar. 1, kami mengevaluasi dengan meningkatkan nilai k, ukuran barang tetangga, digunakan untuk prediksi. Eksperimen kami membandingkan algoritma tradisional, penyaringan kolaboratif berbasis demografi [14], dengan algoritma yang kami usulkan.

Metode CF berbasis demografi meliputi demografi pengguna untuk menemukan kesamaan yang meningkatkan akurasi prediksi. Kami mengamati hasil evaluasi. Ini dapat menunjukkan bahwa jika kita meningkatkan nilai k neighbor, MAE akan menurun dan hasil dari algoritma yang kami usulkan berkinerja lebih baik daripada algoritma lainnya di semua k k tetangga.

1. **Experiment 2**

Kami menguji kinerja dengan meningkatkan jumlah acak dalam data pengujian 5%, 10%, 15%, 20%, 25%, 30%, 35% dan 40% masing-masing, dan mendefinisikan tetangga k adalah sama 5. Itu berarti ada ada 5 film yang mengambil film lingkungan. Hasilnya diilustrasikan pada Gambar. 2. Kami dapat menunjukkan bahwa ketika kami meningkatkan persen dari data pengujian, MAE prediksi meningkat karena peringkat setiap item jarang. Ini mempengaruhi untuk menghitung memilih item untuk tetangga dan menemukan nilai-nilai kesamaan item tetangga, sehingga hasil dari peringkat preferensi prediksi berkurang. Algoritme yang kami usulkan membuat akurasi lebih dari semua.

Hasil dari dua percobaan dapat digunakan untuk mencoba membandingkan tiga algoritma yang ditunjukkan pada Gambar. 3. Kami menentukan jumlah item tetangga sebagai 40. Jumlah film tetangga ini memiliki MAE terendah seperti yang ditunjukkan dalam percobaan 1. Dan kami menghitung dengan menyesuaikan persentase data pengujian masing-masing. Semua algoritma memiliki tren hasil MAE yang sama seperti ketika jumlah data pengujian kecil, MAE rendah. Dalam semua percobaan telah membuat yakin bahwa algoritma yang kami usulkan menawarkan akurasi prediksi lebih dari algoritma lainnya.



**Kesimpulan**

Pemfilteran kolaboratif berbasis item menemukan item yang mirip dengan item target dan memprediksi peringkat preferensi. Secara umum, ia menemukan grup item yang ditutup dengan menggunakan data penilaian setiap item. Dalam penelitian kami, kami mengusulkan algoritma yang mencakup atribut item untuk menghitung kesamaan setiap item, agar efektif pada peringkat preferensi prediksi lebih akurat dalam percobaan ketika kami mengambil nilai yang sama yang diperoleh genre film untuk dihitung. Saat kami memperkirakan peringkat preferensi untuk item target, hasil MAE lebih rendah dari yang lain. Hasil ini menunjukkan bahwa algoritme yang kami usulkan lebih akurat ketika kami bandingkan dengan algoritme Penyaringan Kolaboratif Tradisional dan yang berbasis Demografi dan algoritme yang kami usulkan dapat diimplementasikan ke sistem pemberi rekomendasi saat ini.