



Perbandingan Akurasi dalam Elisitasi Berbasis Peringkat dengan Minat Pengguna

Zen Munawar¹⁾, Novianti Indah Putri²⁾, Yudi Herdiana³⁾, Rita Komalasari⁴⁾

¹⁾⁴⁾Manajemen Informatika, Politeknik LP3I Bandung

Jl. Pahlawan No. 59, Bandung, 40123

²⁾³⁾Teknik Informatika, Fakultas Teknologi Informasi, Universitas Bale Bandung

Jl. R.A.A Wiranatakusumah No. 7, Kec. Baleendah, Kabupaten Bandung, 40375

E-mail : munawarzen@gmail.com¹⁾, noviantiindahputri2021@gmail.com²⁾, ydherdn@gmail.com³⁾, ritakomalasari@plb.ac.id⁴⁾

Abstract

The length of user profiles that must be collected by the recommender system is one of the problems in completing the design of the recommender system. There are two conflicting requirements, firstly the system must collect enough ratings in order to learn the preferences and accuracy of recommendations, secondly collecting more ratings will be a burden to the user and have a negative impact on the experience. The purpose of this study is to find out the impact of subjective user profile length and objectively based accuracy. In this study, three algorithms were applied to offline simulations and conducted online experiments with four recommendation algorithms. By measuring the level of strength of the two contrast strengths that are affected by the number of ratings collected. It was found that the relevance of the recommendations and the burden of the ranking process had a stronger effect on the perceived quality of the user experience. The results of this study were controlled by the user and identified a potentially optimal profile length for an explicit, rank-based elicitation strategy. The value of the length of the profile whose withdrawal reaches 80% of the maximum value of the recall.

Keywords: accuracy, elicitation, rating, interest

Abstrak

Panjangnya profil pengguna yang harus dikumpulkan oleh sistem pemberi rekomendasi menjadi salah satu masalah dalam menyelesaikan rancangan sistem pemberi rekomendasi. Terdapat dua persyaratan yang saling bertolak belakang, yaitu pertama sistem harus mengumpulkan cukup peringkat agar dapat mempelajari preferensi serta akurasi rekomendasi, kedua dengan mengumpulkan lebih banyak peringkat akan menjadi beban untuk pengguna dan berdampak negative atas pengalamannya. Tujuan penelitian ini berusaha mengetahui dampak dari panjangnya profil pengguna secara subjektif dan berbasis akurasi secara objektif. Dalam penelitian ini tiga algoritma diterapkan pada simulasi secara offline serta melakukan eksperimen secara online dengan empat algoritma pemberi rekomendasi. Dengan mengukur tingkat kekuatan dari dua kekuatan kontras yang dipengaruhi oleh jumlah peringkat yang dikumpulkan. Diperoleh relevansi rekomendasi dan beban proses peringkat yang memiliki efek yang lebih kuat pada kualitas yang dirasakan dari pengalaman pengguna. Hasil penelitian ini dilakukan pengendalian oleh pengguna serta mengidentifikasi panjang profil yang berpotensi optimal untuk strategi elisitasi yang eksplisit, berbasis peringkat. Nilai panjang profil yang penarikannya mencapai 80% dari nilai maksimum *recall*nya.

Kata kunci: akurasi, elisitasi, peringkat, minat

1. Pendahuluan

Strategi elisitasi preferensi dapat mempengaruhi "utilitas pengguna baru" (seberapa baik sistem dapat memberikan rekomendasi yang baik kepada pengguna baru yang sedang menjalani proses perolehan) dan "utilitas sistem atau komunitas"

(seberapa baik sistem dapat memberikan rekomendasi yang baik untuk semua pengguna, mengingat apa yang dipelajari dari pengguna baru) [1][2][3]. Selain itu, proses elisitasi mewakili pengalaman awal pengguna dengan pemberi

rekomendasi dan sangat penting untuk membentuk sikapnya terhadap sistem serta proses pengambilan keputusan dan perilakunya (yaitu, apa yang akan dia lakukan dengan rekomendasi). Penelitian sebelumnya sudah menyelidiki apakah ada efek signifikan dalam keakuratan model prediktif [4].

Dunia saat ini tidak lepas dari peran data karena semua dibangun di atas sebuah fondasi data [5]. Sejumlah kriteria desain telah diidentifikasi untuk membuat proses ini lebih efektif baik dari segi utilitas (pengguna baru dan komunitas) dan kualitas interaksi pengguna dengan sistem rekomendasi. Memaksimalkan utilitas dan kualitas penggunaan adalah persyaratan yang saling bertentangan. Jelas, sistem perlu belajar dari pengguna baru dan mengumpulkan preferensi yang cukup untuk menghasilkan rekomendasi yang baik dan memuaskan; tidak mengumpulkan informasi yang cukup dapat mengakibatkan model pengguna yang buruk, yang dapat menyebabkan keakuratan rekomendasi yang terbatas dan pada gilirannya dapat berdampak negatif pada kualitas interaksi pengguna dengan sistem rekomendasi.

Namun, mengharuskan pengguna untuk menghabiskan terlalu banyak waktu dan energi dengan sistem sebelum mereka menerima rekomendasi dapat mengganggu, dan menyebabkan beberapa pengguna menghentikan proses pendaftaran. Saat ini, sejumlah besar data yang dikumpulkan dan dihasilkan setiap hari menawarkan berbagai peluang analitis bagi organisasi untuk mengungkap informasi yang bermanfaat untuk operasinya [6]. Oleh karena itu, pengembang strategi elisitasi harus menghadapi potensi ketegangan desain: meningkatkan utilitas dengan meningkatkan jumlah informasi yang dikumpulkan dari pengguna baru, dan membuat proses elisitasi lancar dari perspektif interaksi pengguna, membatasi kompleksitas dan upaya pengguna selama tugas pendaftaran.

Menemukan kesepakatan yang menyelesaikan ketegangan ini tidak jelas, dan merupakan masalah yang belum terpecahkan dalam penelitian sistem rekomendasi saat ini. Penelitian ini menyelidiki tantangan ini untuk kategori tertentu dari strategi elisitasi, yang dapat dirujuk sebagai eksplisit, berbasis peringkat, dan dikendalikan oleh manusia. Proses elisitasi eksplisit berarti bahwa sistem belajar dari fakta spesifik yang diberikan oleh pengguna baru tentang selera dan preferensi mereka. Dalam proses perolehan berbasis peringkat yang eksplisit, fakta tersebut adalah pendapat pengguna, yaitu peringkat biner atau multi-skala, pada sekumpulan item yang, dalam metode yang dikendalikan manusia, dipilih oleh pengguna itu sendiri. Dalam konteks strategi semacam ini, salah satu ukuran yang mungkin dari upaya pengguna selama proses elisitasi adalah panjang profil - jumlah peringkat yang harus diberikan pengguna baru ke sistem

sebelum mulai menerima rekomendasi. Karenanya fokus pada *tradeoff* yang ada antara memaksimalkan utilitas pengguna dan meminimalkan upaya pemeringkatan. Lebih tepatnya, dapat dilihat pada pertanyaan penelitian berikut: "Manakah dari dua" kekuatan "yang berpotensi kontras yang bergantung pada panjang profil - utilitas pengguna dan upaya pengguna - yang memiliki efek lebih kuat pada kualitas yang dirasakan dari interaksi pengguna?".

Ada dua asumsi implisit dalam pertanyaan penelitian di atas, yang intuitif tetapi tidak selalu dikonfirmasi oleh penelitian sebelumnya. Asumsi pertama adalah panjang profil secara *positif* mempengaruhi utilitas pengguna. Beberapa penelitian sebelumnya menunjukkan bahwa panjang profil pengguna baru berkorelasi positif dengan keakuratan rekomendasi dalam hal utilitas pengguna [1], [7], [8]. Namun, hasil ini tidak dapat dengan mudah digeneralisasikan, karena eksperimen pendukungnya terbatas pada algoritma kolaboratif berbasis item, dan akurasi hanya diukur dalam hal metrik kesalahan: RMSE [7] dan MAE [8].

Selain itu menemukan bahwa korelasi antara panjang profil dan utilitas tidak selalu ada, tetapi bergantung pada strategi perolehan yang diadopsi [9].

Penelitian ini menimbulkan beberapa keraguan pada asumsi umum bahwa profil yang lebih panjang sesuai dengan rekomendasi yang lebih akurat. Rekomendasi bisa juga menimbulkan ketidakpastian yang disebut dengan *certainty factor*. *Certainty factor* adalah metode untuk mengelola ketidakpastian dalam sistem berbasis aturan [10]. Mungkin bertanya-tanya, misalnya, sejauh mana hasil penelitian dapat mengklaim bahwa kejatuhan algoritma pemberi rekomendasi berbasis konten meningkat dengan panjang profil.

Asumsi kedua dalam pertanyaan penelitian di atas adalah bahwa panjang profil berpengaruh negatif terhadap kualitas yang dirasakan dari interaksi pengguna karena beban peringkat. Seperti yang dibahas di bagian selanjutnya, beberapa penelitian telah mengeksplorasi hubungan antara karakteristik strategi elicitation preferensi, upaya pengguna selama pendaftaran, dan interaksi pengguna dengan sistem rekomendasi. Sebelum mengetahui pertanyaan umum untuk semua alasan ini, maka perlu membahas dua pertanyaan penelitian pendahuluan: "Apakah keakuratan algoritma pemberi rekomendasi meningkat seiring dengan panjang profil?" dan "Apakah peningkatan beban pengumpulan peringkat memengaruhi persepsi kualitas interaksi pengguna?".

Tiga pertanyaan penelitian telah ditangani dengan melakukan tiga penelitian utama, yang melibatkan eksperimen *off-line* dan *on-line* yang telah direplikasi dalam kondisi eksperimental yang berbeda, yang melibatkan keseluruhan empat algoritma pemberi rekomendasi dan 960 pengguna.

Dirancang untuk memperoleh preferensi pengguna baru mempengaruhi kualitas yang dirasakan dari interaksi pengguna dengan sistem rekomendasi, dan berdampak pada akurasi keputusan pengguna dan niat untuk kembali [11].

Oleh karena itu sejumlah besar penelitian telah mengeksplorasi proses elisitasi, mencoba memahami bagaimana konstruksi proses preferensi terjadi [12], pertanyaan mana yang akan diajukan kepada pengguna baru [3], item mana yang akan diusulkan [12], dalam bentuk dan memesan [12]. Pembaca dapat merujuk ke [11] untuk ulasan yang lebih lengkap tentang masalah ini. Pada bagian ini, secara singkat menguraikan karya-karya yang lebih relevan dengan konteks penelitian ini.

Dalam hasil *review* [13], membahas *tradeoff* antara akurasi versus usaha, dan menyarankan "meminimalkan munculnya preferensi dalam inisialisasi profil". Argumen untuk pedoman desain ini bersifat teoritis dan empiris. Menurut teori keputusan perilaku [14], pengguna cenderung menerima manfaat langsung dari upaya penghematan atas kepuasan tertunda dari akurasi yang lebih tinggi. Sejumlah karya yang dibahas dalam [9] mendukung prinsip ini (misalnya, [1], [15]), yang juga dikonfirmasi oleh penelitian *online* terbaru yang disajikan di [7], yang menunjukkan bahwa, dalam pemberi rekomendasi berbasis konten, Efektivitas sistem yang dirasakan lebih tinggi terkait dengan pengurangan upaya dalam aktivitas perolehan, diukur dalam hal jumlah penjelajahan sebelum menerima rekomendasi.

Namun, tidak semua penelitian mengkonfirmasi pedoman di atas. Sementara beberapa penelitian menunjukkan bahwa risiko meminta pengguna untuk memberikan terlalu banyak informasi mengganggu mereka [16], atau membuat mereka menyerah dalam proses pendaftaran [9], [17], penelitian lain menunjukkan bahwa pengguna bersedia menghadapi elisitasi yang lebih kompleks jika mereka merasa dihargai dengan rekomendasi yang berguna [9].

Sebagian besar penulis, seperti [3] dan [9], mempertimbangkan metode elisitasi yang dikendalikan sistem dan mengeksplorasi ukuran yang berbeda untuk memilih item untuk dinilai oleh pengguna (misalnya, popularitas, entropi). Oleh karena disarankan bahwa kualitas rekomendasi awal (yaitu, akurasi) dan bukan upaya pengguna harus dipertimbangkan sebagai faktor penentu untuk menilai (dan memilih) strategi elisitasi yang diinginkan.

Hasil serupa diuraikan dalam [18], di mana eksperimen menunjukkan bahwa peringkat yang lebih tinggi tidak selalu berarti lebih banyak upaya yang dirasakan. Namun, temuan ini memiliki motivasi yang berbeda sehubungan dengan [3] dan [9]. Pengguna di [18] menganggap upaya rendah dengan algoritma pemberi rekomendasi kualitas buruk, bahkan dalam kasus proses elisitasi yang

sangat lama, karena mereka merasa perlu untuk memberikan lebih banyak peringkat ke sistem rekomendasi untuk meningkatkan kualitas.

Eksperimen yang dijelaskan dalam [17] mengeksplorasi *tradeoff* desain antara upaya pengguna dan manfaat yang dibawanya baik ke sistem (yang perlu mempelajari tentang pengguna) dan pengguna (yang perlu menerima rekomendasi yang berguna atau meyakinkan) dalam kondisi pengguna yang berbeda. kontrol selama proses elisitasi.

Hasil ini dikonfirmasi oleh sebuah penelitian yang dilaporkan dalam [7], yang menguji efek pada interaksi pengguna metode elisitasi eksplisit vs. implisit. Selain itu, dalam penelitian yang sama, efektivitas sistem dinilai lebih tinggi oleh peserta yang menilai lebih banyak item (karena mereka melihat peningkatan akurasi karena pengetahuan sistem yang lebih luas tentang item tersebut).

Tabel 1. Sekilas Penelitian

Penelitian 1	
Tipe	Simulasi off-line
Ruang lingkup penelitian	Panjang profil vs. utilitas pengguna baru,
Metrik (variabel dependen)	Akurasi (recall dan fallout)
Pertanyaan penelitian	Apakah keakuratan algoritme pemberi rekomendasi meningkat dengan panjang profil?
Algoritma Pengguna	PureSVD, AsySVD, DirectContent 60.000 (disimulasikan)
Panjang profil	5 – 40
Penelitian 2	
Tipe	Eksperimen pengguna online
Ruang lingkup penelitian	Beban proses pemeringkatan
Metrik (variabel dependen)	Kepuasan global
Pertanyaan penelitian	Apakah beban koleksi peringkat yang meningkat mempengaruhi persepsi kualitas interaksi pengguna?
Algoritma Pengguna	TopPop 60 (total)
Panjang profil	5,10
Penelitian 3	
Tipe	Eksperimen pengguna online
Ruang lingkup penelitian	Ketegangan desain antara utilitas dan beban
Metrik (variabel dependen)	Kepuasan global Relevansi yang dirasakan
Pertanyaan penelitian	Manakah dari dua “kekuatan” yang berpotensi kontras yang diciptakan oleh panjang profil - utilitas pengguna dan upaya pengguna - yang memiliki efek lebih kuat pada kualitas interaksi pengguna yang dirasakan?
Algoritma Pengguna	PureSVD, DirectContent 900 (total)
Panjang profil	5,10,20

Sejumlah penelitian telah menganalisis dan membandingkan upaya kognitif yang terkait dengan proses pendaftaran berbasis non-rating (menandai item [1], memunculkan preferensi pengguna pada fitur produk [19], kuis kepribadian [20], dan umpan balik afektif [12]). Dalam percobaan yang dijelaskan

dalam [12], misalnya, metode elisitasi yang kurang menuntut secara kognitif dianggap rendah dalam upaya dan disukai.

Sistem rekomendasi diperlukan karena sebelumnya terdapat kelemahan pada sistem berbasis konten [21]. Panjang profil telah disoroti berkorelasi positif dengan keakuratan rekomendasi, baik dalam hal: utilitas pengguna baru [7], [8], diukur pada sistem rekomendasi kolaboratif dengan metrik kesalahan seperti MAE dan RMSE, utilitas komunitas, seperti memperoleh peringkat untuk film yang tidak memiliki banyak, atau meminta pengguna untuk melakukan pekerjaan yang lebih berharga bagi komunitas (misalnya, memberi *tag* pada konten dan mengespos komentar) [1], [22]. Anehnya, menemukan bahwa korelasi antara panjang profil tidak selalu ada, tetapi bergantung pada strategi elisitasi yang diadopsi [9]

Sebagai penelitian terbaru menemukan bahwa perbaikan dalam MAE dan RMSE tidak selalu merupakan jalan untuk perbaikan dalam pengalaman pengguna [2], akurasi telah diukur dengan menggunakan metrik pengambilan informasi yang lebih luas diadopsi dalam evaluasi sistem rekomendasi komersial [22].

Sebagian besar kumpulan data berisi sejumlah besar item yang tidak diberi peringkat: karena dianggap tidak relevan, mereka kehilangan sebagian kecil dari relevansi positif yang tidak diketahui, dan menyebabkan perkiraan yang terlalu rendah [23]. penelitian sebelumnya menunjukkan bahwa akurasinya adalah salah satu yang terbaik [24] [22].

DirectContent merekomendasikan item yang kontennya mirip dengan konten item yang telah diberi nilai positif oleh pengguna di masa lalu [25].

Misalnya, dalam kasus film, isinya dapat berupa judul, pemerannya, sutradara, genre, dan ringkasannya. *DirectContent* adalah versi sederhana dari algoritma LSA yang dijelaskan dalam [26].

Metodologi pengujian yang diadopsi dalam penelitian ini adalah versi modifikasi dari teknik yang dijelaskan dalam [22].

Kepuasan global adalah indikator bagaimana perasaan pengguna tentang keseluruhan pengalaman dengan sistem, dan merupakan faktor kualitas penting dalam pendekatan yang berpusat pada pengguna untuk evaluasi sistem rekomendasi [6].

Secara khusus, menggunakan algoritma sederhana yang tidak dipersonalisasi (TopPop), yang merekomendasikan item N teratas dengan popularitas tertinggi (jumlah peringkat terbesar)[24].

Relevansi yang dirasakan mengukur seberapa baik pengguna percaya bahwa rekomendasi sesuai dengan minat, preferensi, dan selera, dan, serupa dengan kepuasan global, diakui sebagai faktor kualitas penting dalam kerangka evaluasi yang berpusat pada pengguna [27].

Kepuasan global adalah suatu bentuk "persepsi", yang menunjukkan, dalam terminologi [28].

Selain itu, kelemahannya adalah terbatasnya jumlah atribut yang berpusat pada pengguna yang dipertimbangkan untuk kualitas sistem rekomendasi sehubungan dengan spektrum faktor interaksi pengguna yang diusulkan oleh kerangka kerja yang muncul untuk evaluasi pengalaman pengguna sistem rekomendasi [27][28].

2. Metode Penelitian

Tiga pertanyaan penelitian yang disajikan dalam Bagian 1 (Pendahuluan) telah dieksplorasi dalam tiga (sub) penelitian utama - satu simulasi *off-line* dan dua eksperimen *on-line* - dirangkum dalam Tabel 1.

2.1 Penelitian 1: Akurasi

Penelitian pertama menganalisis keakuratan tiga algoritma pemberi rekomendasi sebagai fungsi dari panjang profil pengguna baru. Secara khusus memfokuskan perhatian pada penarikan kembali (persentase item relevan yang direkomendasikan untuk pengguna) dan kejatuhan (persentase item tidak relevan yang direkomendasikan kepada pengguna).

Penelitian ini mempertimbangkan akurasi vs. panjang profil untuk tiga algoritma: dua algoritma kolaboratif (PureSVD dan AsySVD) dan satu algoritma berbasis konten (*DirectContent*). PureSVD dan AsySVD didasarkan pada faktorisasi matriks.

$$recall = \frac{\# \text{ times the removed 5 stars item is in the list}}{\# \text{ recommendation lists}} \dots \dots \dots (1)$$

Recall pada persamaan (1) didefinisikan sebagai persentase item yang menarik bagi pengguna yang telah direkomendasikan secara efektif oleh sistem.

Pendekatan serupa digunakan untuk mengukur kejatuhan, dengan satu-satunya perbedaan adalah memilih peringkat 1-bintang dari *set probe*, karena dapat secara wajar menyatakan bahwa peringkat ini merujuk pada item yang tidak relevan bagi pengguna. Kejatuhan dihitung sebagai

$$fallout = \frac{\# \text{ times the removed 1 stars item is in the list}}{\# \text{ recommendation lists}} \dots \dots \dots (2)$$

Fallout pada persamaan (2) didefinisikan sebagai persentase item yang tidak menarik bagi pengguna yang telah direkomendasikan secara keliru. Rentang penarikan dan kejatuhan dari 0% hingga 100%.

Algoritma yang ideal harus dapat merekomendasikan semua item yang menarik (yaitu, penarikan sama dengan 100%) dan membuang semua item yang tidak menarik (yaitu, kejatuhan sama dengan 0%).

2.2 Penelitian 2 : Beban

Panjangnya kualitas interaksi yang dirasakan dengan sistem rekomendasi, diukur dalam hal kepuasan global.

Penelitian ini mengukur persepsi kualitas sistem rekomendasi dalam domain film dalam dua kondisi eksperimental yang berbeda. Dalam setiap kondisi eksperimental, dengan menggunakan sistem pemberi rekomendasi yang memiliki kumpulan data yang sama, algoritma yang sama, dan antarmuka pengguna yang sama, tetapi proses penilaian yang berbeda, meminta pengguna untuk memberi peringkat pada jumlah film yang berbeda. Dengan kata lain, dua kondisi eksperimental dicirikan oleh panjang profil yang berbeda (variabel independen), masing-masing 5 dan 10. Karena satu-satunya perbedaan dalam dua kondisi eksperimental adalah upaya pengguna yang objektif (diukur dengan panjang profil), dianggap sebagai algoritma yang tidak dipersonalisasi, yang merekomendasikan daftar item standar yang sama kepada semua orang, terlepas dari profil penggunaannya, sehingga tidak sensitif terhadap panjang profil.

Karena keakuratan rekomendasi yang dihasilkan oleh *TopPop* tidak bergantung pada profil pengguna, satu-satunya gaya yang diukur yang diputar dalam dua kondisi eksperimental profil panjang dan pendek adalah beban peringkat. Oleh karena itu, dalam penelitian ini, mengharapkan kualitas yang dirasakan menurun dengan panjang profil.

2.2.1 Instrumen

PoliRec mendukung pengguna dengan berbagai fungsi yang umum pada layanan persewaan DVD online seperti Netflix dan Lovefilm. Pengguna dapat menelusuri katalog 2137 film, mengambil deskripsi terperinci dari setiap item, memberi peringkat, dan mendapatkan rekomendasi.

2.2.2 Partisipan

Penelitian empiris ini melibatkan 60 subjek, yang dibagi menjadi dua kelompok dengan ukuran yang sama, dan secara acak ditempatkan pada kondisi eksperimen 1 (profil pendek) atau kondisi eksperimen 2 (profil panjang). Karakteristik demografis yang sama dipertahankan di setiap subkelompok: subjek berusia antara 20 dan 50, didistribusikan secara merata ke dalam tiga kategori usia: 20-30, 30-40, 40-50.

Tak satu pun dari mereka yang sebelumnya terpapar sistem penelitian ini, dan tidak ada dari mereka yang memiliki pengetahuan teknis tentang sistem rekomendasi.

2.2.3 Prosedur

Setiap peserta awalnya diminta untuk memberikan informasi pribadinya (usia, jenis kelamin, pendidikan, kebangsaan, dan berapa banyak film yang mereka tonton per bulan). Setelah itu, pengguna diundang untuk menelusuri katalog film menggunakan PoliRec, menilai tingkat apresiasi atau minatnya terhadap film yang ditemui di titik mana pun selama navigasi, menggunakan skala 1 hingga 5 (1 = minat rendah untuk atau apresiasi film; 5 = tinggi). Rekomendasi dibuat setelah peringkat X (X = panjang profil) dikumpulkan.

Semua pengguna diberi tahu bahwa mereka menerima rekomendasi yang dipersonalisasi berdasarkan peringkat masukan mereka, meskipun semua pengguna menerima daftar 5 film populer teratas yang persis sama.

Akhirnya, setiap pengguna diundang untuk mengeksplorasi rekomendasi, untuk menilai relevansi yang dirasakan untuk setiap item yang direkomendasikan pada skala 1 sampai 5 dan untuk menjawab pertanyaan mengenai kepuasan global.

Setiap sesi pengguna berlangsung antara 15 dan 20 menit, dan berlangsung di lingkungan informal, seperti universitas, tempat wawancara, dan tempat wawancara.

2.3 Penelitian 3 : Ketegangan

Dalam setiap penelitian, diukur kualitas sistem rekomendasi yang dirasakan pengguna di domain film dalam tiga kondisi eksperimental yang berbeda. Sama halnya dengan penelitian sebelumnya, ketiga kondisi eksperimen tersebut dicirikan oleh panjang profil yang berbeda (variabel independen), masing-masing 5, 10 dan 20.

Kualitas yang dirasakan pengguna telah dioperasionalkan dalam dua faktor yang dapat diukur (variabel dependen): relevansi yang dirasakan dan kepuasan global (didefinisikan di bagian sebelumnya).

2.3.1 Partisipan

Penelitian empiris secara keseluruhan melibatkan 900 subjek. Dalam masing-masing dari dua penelitian yang direplikasi, penelitian melibatkan 450 subjek yang dibagi dalam tiga kelompok dengan ukuran yang sama, dan secara acak ditugaskan ke kondisi eksperimental 1 (panjang profil = 5 peringkat), kondisi eksperimental 2 (panjang profil = peringkat 10) atau kondisi eksperimental 3 (panjang profil = 20 peringkat). Karakteristik demografis yang sama dipertahankan di setiap subkelompok: subjek berusia antara 20 dan 50, didistribusikan secara merata ke dalam tiga kategori usia: 20-30, 30-40, 40-50. Secara keseluruhan, 52% subjek adalah laki-laki dan 48% perempuan. Tak satu pun dari mereka yang sebelumnya terpapar ke sistem yang digunakan dalam penelitian ini, dan tidak ada dari mereka yang memiliki pengetahuan teknis tentang sistem rekomendasi.

2.3.2 Prosedur

Prosedur yang diadopsi untuk penelitian ini identik dengan yang diadopsi untuk penelitian 2 (Beban), tetapi perekrutan dan pengumpulan data dilakukan oleh tim yang terdiri dari 45 siswa master, yang diorganisir dalam 6 kelompok (2 kelompok per setiap kondisi eksperimen). Mereka terpilih di antara siswa terbaik yang menghadiri kursus "TV Interaktif" di Sekolah Teknik Informasi. Mereka dilatih untuk melakukan pembelajaran, diberikan instruksi tertulis tentang prosedur evaluasi, dan secara teratur diawasi oleh asisten pengajar selama kegiatan mereka. Mereka termotivasi untuk

melakukan evaluasi dengan kemampuan terbaik mereka, karena pekerjaan terus dipantau dan dihitung 20% dari nilai mereka dalam kursus.

3. Hasil dan Pembahasan

Pada bagian ini akan dilakukan analisis dan pembahasan hasil dari ketiga penelitian tersebut.

3.1 Penelitian 1 : Akurasi

Dari analisis dapat mengamati akurasi tersebut dalam syarat penarikan jelas meningkat dengan jumlah peringkat di profil pengguna baru untuk tiga algoritma yang diuji. Sebaliknya, akurasi dalam istilah kejatuhan tidak berubah secara signifikan, dengan pengecualian PureSVD, di mana dalam penelitian ini mengamati peningkatan (yaitu, penurunan) pada nilai kejatuhan. Nilai asimtotik ini, ditemukan untuk setiap algoritma yang diuji, telah memasang *recall* dengan fungsi eksponensial pada persamaan (3).

$$r(l) = a(1 - e^{-bl}) \dots\dots\dots(3)$$

Dimana r adalah penarikan kembali, l adalah jumlah peringkat di profil pengguna baru, dan a dan b adalah dua parameter yang tidak diketahui. Parameter a mewakili nilai asimtotik penarikan, yaitu, penarikan maksimum yang dapat dicapai oleh algoritma dalam hipotesis memiliki jumlah nilai yang sangat besar di profil pengguna. Parameter b mewakili kecepatan di mana *recall* meningkat menuju batas maksimumnya.

Dengan menggunakan kuadrat terkecil, telah menggunakan model eksponensial ke data penarikan. Parameter yang dihasilkan terdaftar dalam Tabel 2. Tabel yang sama menunjukkan nilai panjang profil yang penarikannya mencapai 80% dari nilai maksimum *recall*nya.

Tabel 2. Penelitian 1: Pemasangan *recall* eksponensial

	a	b	Number of ratings at 80% max recall
PureSVD	0,2281	0,1275	13
AsySVD	0,1579	0,1223	13
DirectContent	0,1265	0,2291	7

3.2 Penelitian 2 : Beban

Temuan ini entah bagaimana diharapkan, karena semua pengguna menerima daftar film yang direkomendasikan yang sama, tanpa ada upaya untuk mencocokkan preferensi mereka.

Namun, temuan ini menunjukkan bahwa ada perbedaan yang signifikan dalam kepuasan global antara pengguna profil pendek dan pengguna profil panjang ($p < 0,1$).

3.3 Penelitian 3 : Ketegangan

Penelitian ini menjalankan uji perbandingan berpasangan menggunakan metode *Tukey* untuk membandingkan hasil persepsi relevansi secara lebih analitis. Semua tes dijalankan menggunakan tingkat signifikansi $\alpha = 0,1$.

Saat melihat algoritma PureSVD, relevansi persepsi untuk pengguna baru dengan 10 peringkat

di profil mereka secara signifikan lebih baik daripada relevansi yang dirasakan oleh pengguna dengan peringkat 5 atau 20 ($p < 0,01$). Hal yang sama berlaku untuk algoritma *Direct Content* ($p < 0,1$).

3.4 Pembahasan

Hasil dari dua penelitian pertama memberikan jawaban yang diharapkan entah bagaimana untuk pertanyaan penelitian masing-masing: akurasi rekomendasi dan upaya pengguna yang dirasakan memang meningkat dengan jumlah peringkat yang diperoleh selama pendaftaran pengguna baru.

Lebih khusus lagi, penelitian tentang akurasi (penelitian 1) sebagian mendukung hasil sebelumnya yang menyelidiki hubungan antara akurasi dan panjang profil dalam hal metrik kesalahan (MAE dan RMSE) untuk dua sistem rekomendasi berbasis item. Penelitian ini telah memberikan bukti empiris bahwa ingatan meningkat dengan panjang profil, dan korelasi ini ada untuk algoritma yang berbeda. Namun, hal yang sama tidak terjadi untuk kejatuhan, yang tidak berkorelasi dengan panjang profil.

Menarik untuk membandingkan hasil penelitian 2 dengan temuan penelitian 3. Menurut penelitian 2, upaya pengguna yang meningkat selama proses pendaftaran berdampak negatif terhadap kualitas interaksi pengguna yang dirasakan, dalam hal kepuasan global, jika tambahan beban tidak dikompensasikan dengan peningkatan utilitas (yaitu, relevansi rekomendasi yang lebih baik). Efek ini terlihat dalam eksperimen penelitian karena kualitas (relevansi) yang rendah dari rekomendasi yang diberikan oleh algoritma yang tidak dipersonalisasi yang di adopsi. Pengguna diberi sedikit penghargaan oleh rekomendasi yang berguna, terlepas dari jumlah peringkat yang lebih tinggi, dan mereka lebih rentan untuk merasakan beban tambahan.

Fenomena yang sama tidak terjadi di penelitian 3, di mana pengguna menerima rekomendasi yang berkualitas baik (relevan). Jika proses pemeringkatan yang lebih menuntut diimbangi dengan rekomendasi yang jauh lebih baik, kepuasan global tidak terpengaruh secara negatif oleh peningkatan upaya. Seolah-olah dua kekuatan kontras (akurasi dan upaya pengguna) yang dihasilkan oleh panjang profil pada kualitas interaksi pengguna saling mengimbangi. Efek yang berpotensi positif dari peningkatan akurasi yang dihasilkan dari profil yang lebih panjang terkikis oleh beban proses pemeringkatan yang lebih menuntut, tetapi efek ini tidak cukup kuat untuk menurunkan opini global pengguna terhadap sistem pemberi rekomendasi.

Namun, ketika membandingkan nilai indikator yang berbeda dari kualitas interaksi pengguna yang dirasakan relevansi yang dirasakan dalam kondisi panjang profil yang berbeda. Dengan peringkat 5 dan 10, dapat mengamati bahwa relevansi yang dirasakan meningkat dengan jumlah peringkat - ini diharapkan, karena rekomendasi yang lebih akurat.

Namun, saat membandingkan panjang profil dengan peringkat 10 dan 20, penelitian ini mengamati perilaku yang mengejutkan: relevansi yang dirasakan menurun. Temuan ini menegaskan intuisi bahwa, karena relevansi rekomendasi tidak meningkat tanpa batas dengan panjang profil, harus ada jumlah peringkat maksimum yang dapat diperoleh dari pengguna baru tanpa gaya negatif yang disebabkan oleh peningkatan beban mengatasi gaya positif relevansi. Hasilnya juga koheren dengan temuan penelitian 1, yang menunjukkan (lihat Tabel 2) bahwa nilai maksimum untuk recall mendekati 10 peringkat untuk algoritma yang dipertimbangkan dalam penelitian 3, dan oleh karena itu menunjukkan bahwa panjang profil lebih dari 10 tidak bertambah relevansi yang dirasakan dari rekomendasi.

Apakah aspek objektif tertentu dari interaksi dengan sistem terdaftar dengan pengguna sama sekali; relevansi yang dirasakan adalah bentuk "evaluasi", yang menunjukkan apakah aspek yang dipersepsikan memiliki nilai pribadi bagi pengguna.

Hipotesis dapat berupa ketika penilaian pengguna terkait dengan persepsi yang diwakili oleh kepuasan global semua aspek pengalaman yang dirasakan dengan sistem diintegrasikan ke dalam perspektif yang lebih holistik, dan saling memberi kompensasi, kecuali beberapa kekuatan secara signifikan lebih kuat atau lebih lemah dari pada yang lain, seperti yang terjadi dalam penelitian 2. Ketika penilaian pengguna bergerak ke arah refleksi yang lebih berorientasi pada nilai seperti yang diungkapkan oleh relevansi yang dirasakan mekanisme yang ditunjukkan, mulai berlaku, dan pengguna cenderung untuk fokus pada manfaat langsung dari upaya penghematan atas manfaat yang kurang (bagi mereka) terukur dari akurasi yang lebih tinggi.

4. Kesimpulan

Penelitian ini menjelaskan beberapa masalah yang belum terselesaikan dalam desain sistem rekomendasi: bagaimana sistem dapat mengumpulkan informasi (peringkat) yang "cukup" dari pengguna, untuk mempelajari preferensinya dan meningkatkan kualitas rekomendasi tanpa menambahkan kelebihan beban pada pengguna, yang dapat menghilangkan manfaat yang dirasakan dari rekomendasi hasil personalisasi yang baik? Hasil penelitian telah menyelidiki masalah ini dari berbagai perspektif, dalam konteks teknik elisitasi yang dikendalikan pengguna berbasis rating, melalui serangkaian penelitian *offline* dan *online* yang luas. Validitas temuan terbatas pada algoritma aktual dan kondisi eksperimental yang dipertimbangkan. Namun, dalam bidang di mana pekerjaan empiris sangat kompleks dan menuntut sumber daya, penelitian ini mewakili penelitian yang luas dan terartikulasi yang menjembatani evaluasi online yang berpusat pada pengguna dengan evaluasi

offline, dan memberikan kontribusi untuk penelitian sistem rekomendasi dan praktik desain.

Dari perspektif desain, temuan (Penelitian 1 dan 3) menunjukkan bahwa jumlah peringkat optimal dalam domain film adalah antara peringkat 5 dan 20 (lebih mungkin peringkat 10): dalam rentang inilah kekuatan kontras yang disebabkan oleh panjang profil mencapai keseimbangan yang lebih baik, dari perspektif kualitas interaksi pengguna. Hasil ini dapat disaring ke dalam heuristik: "10 peringkat sudah cukup", yang dapat membantu *desainer* untuk memprioritaskan keputusan desain dan menunjukkan bahwa tidak ada kebutuhan nyata untuk sistem bangunan yang mengumpulkan profil yang sangat panjang. Dari perspektif teoritis, temuan pada penelitian 1 menunjukkan bahwa tidak dapat diberikan begitu saja bahwa kualitas rekomendasi yang obyektif (yaitu, diukur secara statistik) meningkat dengan bertambahnya panjang profil. Telah diberikan bukti empiris, untuk algoritma yang berbeda, bahwa ada korelasi positif antara penarikan kembali dan panjang profil. Namun, dalam percobaan fenomena yang sama tidak terjadi untuk kejatuhan, yang tidak berkorelasi dengan panjang profil; ini menyimpulkan bahwa kesalahan rekomendasi hanya bergantung pada algoritma, dan sebagian besar tidak bergantung pada jumlah peringkat yang dikumpulkan dari pengguna baru. Akhirnya, temuan penelitian 3 mengkonfirmasi dan memperluas beberapa hasil penelitian sebelumnya: metrik akurasi mengukur kekuatan interaksi pengguna yang "lemah", yang kurang penting daripada yang diharapkan dalam meningkatkan persepsi pengguna tentang sebuah kualitas pemberi rekomendasi.

Saran untuk penelitian selanjutnya agar akurasinya lebih tinggi penelitian yang dilakukan sebaiknya lebih dari tiga.

5. Daftar Pustaka

- [1] S. Drenner, S. Sen, and L. Terveen, 2008, "Crafting The Initial User Experience to Achieve Community Goals," in RecSys'08: Proceedings of the 2008 ACM Conference on Recommender Systems, no. May, pp. 187–194.
- [2] J. A. Konstan and J. Riedl, 2012, "Recommender systems: From algorithms to user experience," User Model. User-adapt. Interact., vol. 22, no. 1–2, pp. 101–123.
- [3] A. M. Rashid, 2002 "Getting to Know You: Learning New User Preferences in Recommender Systems," in 7th International Conference on Intelligent User Interfaces (IUI '02), pp. 127–134.
- [4] Z. Munawar, Dec. 2017, "Penggunaan Profil Media Sosial Untuk Memprediksi Kepribadian," Temat. - J. Teknol. Inf. dan Komun., vol. 4, no. 2 SE-Articles, pp. 18–37.
- [5] Z. Munawar, B. Siswoyo, and N. S. Herman, 2017, "Machine learning approach for analysis of social media," ADRI Int. Journal. Information. Technol., vol. 1, pp. 5–8.
- [6] N. I. Munawar, Zen and Putri, 2020, "Keamanan Jaringan Komputer Pada Era Big Data," J-SIKA| J. Sist. Inf. Karya Anak Bangsa, vol. 02, no. 01, pp. 14–20.
- [7] N. Golbandi, Y. Koren, R. Lempel, H. D, 2010. Management, and D. Applications, "On bootstrapping recommender systems-golbandi," pp. 1805–1808.
- [8] S. Berkovsky, Y. Eytani, T. Kuflik, and F. Ricci, 2007,

- “Enhancing privacy and preserving accuracy of a distributed collaborative filtering,” in *RecSys’07: Proceedings of the 2007 ACM Conference on Recommender Systems*, pp. 97–104.
- [9] A. M. Rashid, G. Karypis, and J. Riedl, 2008, “Learning preferences of new users in recommender systems,” *ACM SIGKDD Explor. Newsl.*, vol. 10, no. 2, pp. 90–100.
- [10] N. I. Putri, 2018, “Sistem pakar diagnosa tingkat kecanduan gadget pada remaja menggunakan metode Certainty Factor,” UIN Sunan Gunung Djati Bandung.
- [11] L. Chen and P. Pu, *Interaction design guidelines on critiquing-based recommender systems*, vol. 19, no. 3. Springer, 2009.
- [12] A. Pommeranz, J. Broekens, P. Wiggers, W. P. Brinkman, and C. M. Jonker, 2012, “Designing interfaces for explicit preference elicitation: A user-centered investigation of preference representation and elicitation process,” *User Model. User-adapt. Interact.*, vol. 22, no. 4–5, pp. 357–397.
- [13] P. Pu, L. Chen, and R. Hu, 2012, “Evaluating recommender systems from the user’s perspective: Survey of the state of the art,” *User Model. User-adapt. Interact.*, vol. 22, no. 4–5, pp. 317–355.
- [14] G. Häubl and V. Trifts, 2000, “Consumer decision making in online shopping environments: The effects of interactive decision aids,” *Mark. Sci.*, vol. 19, no. 1, pp. 4–21.
- [15] N. Jones and P. Pu, 2007, “[cold start] User Technology Adoption Issues in Recommender Systems,” in *Networking and Electronic Commerce Research Conference*, no. January, pp. 379–394.
- [16] G. Lekakos and G. M. Giaglis, 2007, “A hybrid approach for improving predictive accuracy of collaborative filtering algorithms,” *User Model. User-adapt. Interact.*, vol. 17, no. 1–2, pp. 5–40.
- [17] S. M. McNee, J. A. Konstan, J. Riedl, and S. K. Lam, 2014, “Interfaces for Eliciting New User Preferences in Recommender Systems,” in *User Modeling 2003, 9th International Conference, 2003*, vol. 2702, no. June 2014, pp. 163–167.
- [18] K. Swearingen and R. Sinha, 2002, “Interaction design for recommender systems,” in *Designing Interactive Systems, 2002*, no. March 2002, pp. 1–10.
- [19] A. Adomavicius, G. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions,” in *IEEE Transactions on Knowledge and Data Engineering*, 2005, vol. 17, no. 6, pp. 734–749.
- [20] R. Hu and P. Pu, 2009, “A comparative user study on rating vs. personality quiz based preference elicitation methods,” in *International Conference on Intelligent User Interfaces, Proceedings IUI*, pp. 367–371.
- [21] Z. Munawar, N. Suryana, Z. B. Sa’aya, and Y. Herdiana, “Framework With An Approach To The User As An Evaluation For The Recommender Systems,” in *2020 Fifth International Conference on Informatics and Computing (ICIC), 2020*, pp. 1–5.
- [22] P. Cremonesi, Y. Koren, and R. Turrin, 2010 “Performance of recommender algorithms on top-N recommendation tasks,” in *RecSys’10 - Proceedings of the 4th ACM Conference on Recommender Systems*, , pp. 39–46.
- [23] A. Bellogín, P. Castells, and I. Cantador, 2011, “Precision-oriented evaluation of recommender systems: An algorithmic comparison,” in *RecSys’11 - Proceedings of the 5th ACM Conference on Recommender Systems*, pp. 333–336.
- [24] P. Cremonesi, F. Garzotto, and R. Turrin, 2012, “Investigating the persuasion potential of recommender systems from a quality perspective: An empirical study,” *ACM Trans. Interact. Intell. Syst.*, vol. 2, no. 2, pp. 1–41.,
- [25] P. Lops, M. De Gemmis, and G. Semeraro, 2011, *Recommender Systems Handbook* ISBN 978-0-387-85819-7 e-ISBN 978-0-387-85820-3 DOI 10.1007/978-0-387-85820-3, London; Springer Science+Business Media, LLC, pp. 1-848. https://www.cse.iitk.ac.in/users/nsrivast/HCC/Recommend_er_systems_handbook.pdf.
- [26] R. Bambini, P. Cremonesi, and R. Turrin, . 2011, *Recommender Systems Handbook*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-85820-3_9. pp 299-331 https://link.springer.com/chapter/10.1007/978-0-387-85820-3_9.
- [27] R. Hu and P. Pu, 2011, “Enhancing collaborative filtering systems with personality information,” in *RecSys’11 - Proceedings of the 5th ACM Conference on Recommender Systems*, pp. 197–204.
- [28] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, and C. Newell, 2012, “Explaining the user experience of recommender systems,” *User Model. User-adapt. Interact.*, vol. 22, no. 4–5, pp. 441–504.