

Penerapan Content-Boosted Collaborative Filtering untuk Meningkatkan Kemampuan Sistem Rekomendasi Penyedia Jasa Acara Pernikahan

Rahman Indra Kesuma¹, Amirul Iqbal²

^{1,2}Program Studi Teknik Informatika, Jurusan Teknologi Produksi dan Industri,
Institut Teknologi Sumatera

Jl. Terusan Ryacudu, Way Hui, Kec. Jati Agung, Lampung Selatan, 35365

¹rahman.indra@if.itera.ac.id, ²amirul.iqbal@if.itera.ac.id

Abstract

The changes in lifestyle of the global society in the era of digital world development have made the smartphone technology penetration to rise continually. This condition can increase business opportunities, especially e-commerce activities that utilize technology and the internet in terms of promotions and transactions. The efficiency and effectiveness is an interesting focus that is discussed in this issue. For example, in services or products searching for a wedding where many customers still feel difficult and need a long time to find the desired things. The existence of a recommendation system also has not been able to help, especially for users who are newly registered to the system. This is because most of them will provide recommendations based on a history of user activity. Therefore, this study applies the content-boosted collaborative filtering (CBCF) method to improve the ability of the recommendation system in providing recommendations for weddings, especially for a new user. The obtained results are then compared with two commonly used methods, content-based recommendations (CB) and collaborative filtering (CF). Based on the experimental results, it can be concluded that CBCF can maintain the quality of good recommendations for long registered users with an accuracy of 84% and also can provide recommendations for new users with an accuracy of 54% which is cannot be solved by CB or CF methods.

Key Word: digital businesses, wedding vendors/organizers, recommendation system, content-boosted collaborative filtering

Abstrak

Perubahan pola kehidupan masyarakat global pada era perkembangan dunia digital membuat penetrasi dari teknologi telepon pintar terus menaik. Kondisi ini dapat meningkatkan kesempatan bisnis khususnya kegiatan jual beli yang memanfaatkan teknologi dan internet dalam hal promosi dan transaksi. Efisiensi dan efektifitas proses menjadi fokus yang terus menarik dibahas dalam hal ini. Sebagai contoh, pada pencarian layanan atau produk untuk pernikahan yang mana banyak pelanggan masih merasakan kesulitan dan membutuhkan waktu yang lama untuk mencari sesuatu yang diinginkannya. Keberadaan sistem rekomendasi juga belum bisa membantu terlebih bagi pengguna yang baru terdaftar pada sistem. Hal ini dikarenakan kebanyakan sistem akan memberikan rekomendasi berdasarkan rekam jejak aktifitas pengguna. Maka itu, pada penelitian ini diusulkan penerapan metode content-boosted collaborative filtering (CBCF) untuk meningkatkan kemampuan sistem rekomendasi dalam pemberian rekomendasi untuk acara pernikahan, khususnya pada pengguna baru. Hasil yang diperoleh selanjutnya dibandingkan dengan dua metode yang umum digunakan yaitu content based recommendation (CB) dan collaborative filtering (CF). Berdasarkan hasil penelitian yang diperoleh, dapat disimpulkan bahwa CBCF dapat mempertahankan kualitas pemberian rekomendasi yang baik untuk pengguna lama dengan akurasi sebesar 84% serta mampu memberikan rekomendasi untuk pengguna baru dengan akurasi 54% yang mana kondisi ini tidak bisa diselesaikan oleh metode CB ataupun CF.

Kata Kunci: bisnis digital, penyedia jasa acara pernikahan, sistem rekomendasi, content-boosted collaborative filtering

I. PENDAHULUAN

Perkembangan dunia yang pesat ke arah teknologi digital mendorong perubahan dalam segala aspek kehidupan masyarakat. Hal ini telah dibuktikan bahwa di Indonesia pengguna teknologi ini cukup tinggi yang mana dilihat dari kepemilikan telepon pintar di daerah perkotaan yang sebesar 70,96% serta di daerah perdesaan sebesar 42,06% [1]. Fenomena ini juga dirasakan pada masyarakat dunia, yang mana penetrasi teknologi ini sebesar 43% dari populasi dunia dan diprediksi akan terus bertambah hingga menjadi 61% pada tahun 2025 [2]. Berdasarkan kondisi tersebut seharusnya bisa dimanfaatkan bagi banyak pihak untuk menjangkau sejumlah masyarakat untuk kepentingan tertentu. Salah satunya adalah pelaku bisnis baik itu dari usaha mikro, kecil, menengah hingga besar.

Fenomena tersebut membuat pertumbuhan *e-commerce* menjadi lebih baik karena kemudahan akses dengan menggunakan perangkat telepon pintar [3]. Istilah *e-commerce* yang merujuk pada kegiatan atau transaksi bisnis menggunakan media teknologi informasi serta pemanfaatan internet saat ini terus berkembang [4]. Perkembangan ini dapat terlihat pada banyaknya pengguna internet yang melakukan transaksi *online* melalui *e-commerce*, yaitu sejumlah 1,7 miliar pengguna dari 4 triliun pengguna internet di dunia [5]. penggunaan bisnis *online* yang terus berkembang banyak telah dirasakan untuk memenuhi beberapa aspek kebutuhan masyarakat. Salah satunya yaitu untuk kepentingan persiapan acara pernikahan bagi calon pengantin.

Kebanyakan calon pengantin menginginkan acara yang terbaik pada hari bahagiannya. Pertimbangan yang cukup banyak membuat pilihan yang tersedia semakin banyak, hal ini mengakibatkan banyak calon pengantin yang tidak memiliki cukup waktu untuk melakukan survei terhadap penyedia jasa pernikahan tersebut [6]. Cukup banyaknya website pernikahan yang dapat menyediakan jasa dan produk untuk acara, ternyata belum dapat memudahkan calon pengantin untuk melakukan pemilihan, karena membutuhkan waktu yang cukup untuk membuka dan mencari produk atau jasa yang sesuai dengan keinginan dari calon pengantin [7]. Maka dari itu, perlunya sebuah terobosan baru bukan hanya menyediakan sebuah website ataupun media untuk menggabungkan dari beragam website penyedia jasa untuk keperluan pernikahan. Terobosan yang coba dihadirkan pada penelitian ini yaitu penerapan konsep sistem rekomendasi untuk mempermudah pencarian dan promosi produk dan jasa dari acara pernikahan.

Sistem rekomendasi adalah suatu sistem yang memiliki tujuan untuk memberikan rekomendasi produk barang atau jasa baru yang berpotensi digunakan ataupun diakses oleh seorang pengguna (calon pembeli) [8]. Teknik kompleks mendasar yang dapat digunakan dalam sistem rekomendasi adalah *Content-Based* (CB), *Collaborative Filtering* (CF) serta *Hybrid Recommender*. Metode CB berfokus pada pemberian rekomendasi berdasarkan kesamaan deskripsi antara produk yang selanjutnya dihubungkan dengan pengalaman pribadi pengguna pada suatu produk [9]. Sedangkan metode CF menggunakan hubungan antar pengguna ataupun hubungan antar produk dalam pemberian rekomendasi, berdasarkan pengalaman transaksi yang sudah terjadi sebelumnya [9]. Selain dua pendekatan metode tersebut, terdapat pendekatan hibrida yang berfokus untuk menggabungkan dua atau lebih metode rekomendasi untuk memperoleh keunggulan dari tiap-tiap algoritma dan menekan kelemahan dari algoritma tersebut. Pada penelitian ini digunakan metode *Content-Boosted Collaborative Filtering* (CBCF) yang memanfaatkan keunggulan dari CB dan CF untuk saling menutupi kekurangan dari setiap metode.

Banyak penelitian sebelumnya yang mengangkat topik sistem rekomendasi yang digunakan untuk meningkatkan promosi ataupun penjualan produk dari suatu bisnis online. Pendekatan metode *content-based filtering* digunakan dalam penelitian [10], yang mana pemberian rekomendasi dengan menggunakan eksplorasi informasi dari profil pengguna, deksripsi produk, serta hal-hal lain yang berhubungan dengan transaksi yang terjadi. Metode ini melakukan analisis kemiripan profil pengguna dengan vektor komponen penyusun produk, yang selanjutnya dapat memprediksi produk yang akan disukai oleh pengguna, serta layak untuk direkomendasikan kepada pengguna. Namun, terdapat kekurangan yang utama dari metode ini yaitu tidak mampu memberikan rekomendasi terhadap produk baru ataupun produk yang belum pernah dilihat oleh pengguna.

Pendekatan metode lainnya adalah *collaborative filtering*, yang mana pernah diterapkan pada penelitian [11] untuk membuat mekanisme rekomendasi dalam Netflix dengan kumpulan data yang mengandung 17 ribu film dan lebih dari 500 ribu pengguna. Metode *item-based collaborative filtering* digunakan dengan melakukan pencarian hubungan kemiripan antara produk dengan mempertimbangkan rekam jejak transaksi antar produk terhadap pengguna yang ada. Metode ini dianggap metode yang mudah untuk diimplemetasikan dan efisien dalam menghasilkan solusi. Namun, terdapat banyak tantangan dan keterbatasan yang terdapat dalam metode ini, yaitu sering dikenal dengan istilah *sparsity* dan *scalability*. *Sparsity* merupakan suatu kondisi dimana sistem rekomendasi dihadapkan pada produk yang tidak populer ataupun baru, maka produk ini kemungkinan besar tidak akan pernah masuk dalam produk yang akan direkomendasikan. Sedangkan *scalability*, merupakan kondisi yang mana komputasi yang akan terus bertambah karena metode ini memiliki ketergantungan terhadap hubungan antara produk dan pelanggan. Sehingga bertambahnya jumlah produk dan pelanggan, akan mengakibatkan biaya komputasi yang juga bertambah.

Permasalahan yang muncul dari dua metode sebelumnya, memberikan tantangan dalam optimalisasi kinerja dari sistem rekomendasi. Optimalisasi dari metode yang ada untuk sistem rekomendasi biasanya menggunakan pendekatan metode hibrida. Metode *Content-Boosted Collaborative Filtering* (CBCF) dalam penelitian [9] merupakan salah satu metode yang menerapkan pendekatan tersebut. Metode ini pada intinya menggunakan dua metode rekomendasi yaitu *content-based* (CB) serta *collaborative filtering* (CF). Proses diawali dengan menerapkan CB untuk mengubah matriks penilaian pengguna terhadap produk, yang tidak lengkap menjadi matriks penilaian yang lengkap. Selanjutnya, menggunakan CF untuk melakukan rekomendasi. Hasil penelitian memperoleh hasil bahwa metode hibrida ini bisa menyelesaikan permasalahan *sparsity* yang dihadapi pada dua metode sebelumnya, namun tetap masih perlunya perbaikan untuk bisa mengoptimalkan kinerja metode hibrida ini. Salah satunya dengan menggunakan metode prediksi yang cukup baik untuk membangun matriks penilaian yang lengkap dari hubungan pengguna dan produk.

II. METODOLOGI PENELITIAN

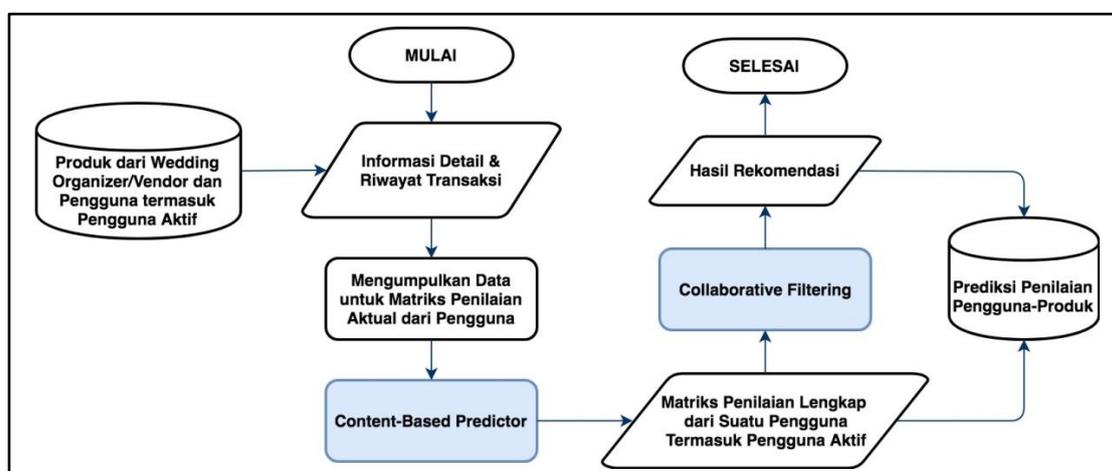
A. Deskripsi Domain

Pada penelitian ini, penerapan metode CBCF berada pada domain pemberian nilai rating dari calon pengantin terhadap produk/jasa dekorasi yang disediakan oleh enam wedding organizer/vendor. Kumpulan data yang digunakan sebagai data latih maupun data uji diperoleh dari survei yang telah dilakukan terhadap 14 pengguna dan 25 produk/jasa dekorasi dari beragam vendor tersebut. Sistem akan membangun informasi profil pada setiap pengguna yang memiliki beragam karakteristik, dengan berdasarkan pada parameter umur, pekerjaan, suku, tema pernikahan, jenis musik kesukaan, warna kesukaan, dana pernikahan, jumlah tamu undangan serta status penggunaan wedding organizer. Sedangkan produk dibangun informasi profilnya oleh sistem dengan menggunakan parameter deskripsi dan harga produk.

Pada tahap proses pembelajaran dan pengujian terhadap algoritme CBCF pada 14 pengguna yang telah memberikan penilaian kepada produk/jasa dekorasi, disimulasikan bahwa 4 pengguna diperlakukan sebagai pengguna yang baru terdaftar pada sistem (data uji) sedangkan 10 pengguna sisanya diperlakukan sebagai pengguna yang telah lama terdaftar pada sistem dan telah memiliki riwayat penilaian. Selain itu, pasangan penilaian dari 10 pengguna terhadap 25 produk/jasa dekorasi, hanya digunakan 150 penilaian yang dipilih secara acak untuk membentuk matriks penilaian sebenarnya dari seorang pengguna (data latih).

B. Deskripsi Sistem Rekomendasi

Garis besar dari proses rekomendasi dengan menggunakan metode CBCF yang diterapkan pada penelitian ini dapat diamati pada Gambar 1. Proses diawali dengan mengumpulkan informasi detail terkait dengan konten produk/jasa dari wedding organizer/vendor, serta informasi detail yang berkaitan dengan pengguna. Berikutnya, riwayat transaksi pemberian penilaian dari pengguna terhadap produk dikumpulkan untuk membentuk matriks penilaian aktual dari pengguna. Selanjutnya, proses dimulai dengan penerapan metode rekomendasi yang menggunakan pendekatan hibrida dengan menerapkan dua metode umum dalam rekomendasi, yaitu CB dan CF [9].



Gambar 1. Alur Proses Rekomendasi dari CBCF

Pada dasarnya, penerapan metode CBCF menggunakan metode CB pada awal proses untuk mengubah matriks penilaian yang tidak lengkap dari pengguna menjadi matriks penilaian yang lengkap, yang dilanjutkan dengan penggunaan metode CF untuk membuat rekomendasi. Tahapan yang dilakukan dalam penerapan metode ini adalah sebagai berikut:

- Membentuk matrik penilaian pengguna, untuk setiap pengguna yang terdaftar. Matriks penilaian mengandung nilai $v_{u,i}$ yang memiliki makna penilaian dari pengguna u terhadap produk i , dengan mempertimbangkan formula berikut

$$v_{u,i} = \begin{cases} r_{u,i} & \text{: Jika pengguna } u \text{ telah menilai produk } i \\ c_{u,i} & \text{: Selain kondisi tersebut.} \end{cases} \quad (1)$$

dimana $r_{u,i}$ merupakan data sebenarnya dari penilaian pengguna u terhadap produk i , sedangkan $c_{u,i}$ merupakan prediksi penilaian dengan menggunakan metode CB.

- Metode CB akan memberikan perlakuan prediksi sebagai permasalahan pengkategorian teks. Informasi dari produk digunakan sebagai teks dokumen dan penilaian dari pengguna u digunakan sebagai label kelas. Konsep pengklasifikasian teks Naïve Bayes digunakan dalam dua kondisi klasifikasi berikut:
 - Riwayat penilaian pengguna terhadap produk akan digunakan untuk memprediksi nilai rating pada produk yang belum pernah dinilai oleh suatu pengguna. Proses dilakukan dengan mempertimbangkan nilai similaritas dari deskripsi produk antara produk yang belum dinilai dan produk yang sudah dinilai.
 - Jika metode CB dihadapkan pada kondisi dimana pengguna aktif merupakan pengguna baru, maka riwayat penilaian dari pengguna lain akan digunakan untuk memprediksi penilaian dari suatu produk yang akan diberikan oleh pengguna baru tersebut. Proses dilakukan dengan memperhatikan similaritas dari profil pengguna baru dengan profil pengguna yang telah terdaftar dan memberikan penilaian pada sistem.
- Setelah matriks penilaian dari setiap pengguna kepada setiap produk sudah berhasil terbentuk (matriks penilaian lengkap), proses rekomendasi masuk ke tahapan metode CF. Proses ini dimulai dengan pemberian bobot pada semua pengguna dengan mempertimbangkan similaritas dengan pengguna lainnya.

$$p_{a,u} = \frac{\sum_{i=1}^m (v_{a,i} - \bar{v}_a) \times (v_{u,i} - \bar{v}_u)}{\sqrt{\sum_{i=1}^m (v_{a,i} - \bar{v}_a)^2 \times (v_{u,i} - \bar{v}_u)^2}} \quad (2)$$

dimana $v_{a,i}$ merupakan data penilaian yang diberikan pada produk i oleh pengguna a . Selanjutnya \bar{v}_a merupakan rata-rata penilaian yang diberikan oleh pengguna a . Sedangkan m merupakan jumlah total dari produk yang digunakan sebagai perbandingan antar pengguna.

- Selanjutnya menghitung *Harmonic Mean Weighting* ($hm_{i,j}$) serta *Hybrid Correlation Weight* ($hw_{a,u}$) untuk menghitung bobot korelasi antar pengguna, dengan menggunakan formula berikut

$$hm_{i,j} = \frac{2m_i m_j}{m_i + m_j}, \text{ dimana } m_u = \begin{cases} \frac{n_u}{\beta} & \text{: jika } n_u < \beta \\ 1 & \text{: jika selainnya} \end{cases} \quad (3)$$

$$hw_{a,u} = hm_{a,u} + sg_{a,u}, \text{ dimana } sg_{a,u} = \begin{cases} \frac{q}{\beta} & \text{: jika } q < \beta \\ 1 & \text{: jika selainnya} \end{cases} \quad (4)$$

dimana n_u merupakan jumlah produk yang sudah dinilai oleh pengguna u , β merupakan parameter jumlah maksimum produk yang harus sudah dinilai sebagai bahan pertimbangan (pada penelitian ini digunakan nilai 15), sedangkan q merupakan jumlah produk yang sama dan telah dinilai baik oleh pengguna a maupun pengguna u .

- Berikutnya adalah menghitung *Self Weighting* (sw_a) yang digunakan sebagai bobot untuk meningkatkan tingkat kepercayaan terhadap nilai prediksi hasil dari pemrosesan metode CB untuk suatu pengguna.

$$sw_a = \begin{cases} \frac{n_a}{\beta} \times max & \text{: jika } n_a < \beta \\ max & \text{: jika selainnya} \end{cases} \quad (5)$$

dengan mempertimbangkan n_a adalah jumlah produk yang sudah dinilai oleh seorang pengguna, β merupakan parameter yang sama seperti pada persamaan (3) serta max yang merupakan nilai maksimum bobot untuk tingkat kepercayaan tertinggi dari nilai prediksi.

- Setelah memperoleh semua nilai yang dibutuhkan, maka perhitungan yang dilakukan terakhir adalah menghitung nilai prediksi ($p_{a,i}$) dari seorang pengguna a terhadap produk i .

$$p_{a,i} = \begin{cases} \bar{v}_a + \frac{sw_a \times (c_{a,i} - \bar{v}_a) + \sum_{u=1, u \neq a}^n hw_{a,u} \times p_{a,u} \times (v_{u,i} - \bar{v}_u)}{sw_a + \sum_{u=1, u \neq a}^n hw_{a,u} \times p_{a,u}} & \text{: Jika } a \text{ merupakan pengguna baru} \\ \bar{v}_a + \frac{\sum_{u=1, u \neq a}^n p_{a,u} \times (v_{u,i} - \bar{v}_u)}{\sum_{u=1, u \neq a}^n p_{a,u}} & \text{: Selain itu} \end{cases} \quad (6)$$

Mempertimbangkan semua parameter yang telah dipaparkan pada langkah sebelumnya, yang mana $c_{a,i}$ merupakan nilai prediksi *content-based* dari pengguna a terhadap produk i , $v_{u,i}$ merupakan nilai matriks penilaian dari pengguna u terhadap produk i , \bar{v}_u merupakan nilai rata-rata penilaian dari semua produk yang dinilai oleh pengguna u , n yang berisi jumlah tetangga terdekat dari pengguna a , serta nilai $p_{a,u}$, $hw_{a,u}$ serta sw_a yang merupakan hasil dari persamaan (2), (4) dan (5) secara berurutan.

- Pada akhirnya untuk setiap pengguna a , ambilah sejumlah produk i yang belum pernah dilakukan penilaian, dengan mempertimbangkan nilai $p_{a,i}$ terbesar dari persamaan (6), untuk dilakukan rekomendasi ke pengguna a .

C. Metode Pemandangan

Metode yang diterapkan pada sistem rekomendasi di penelitian ini selanjutnya dibandingkan dengan dua metode yang umum digunakan dalam sistem rekomendasi. Dua metode tersebut adalah *Collaborative Filtering* (CF) berdasarkan penelitian [12] serta *Content Based Predictor* (CB) dengan penerapan pengkategorian teks [13]. Metode CF bekerja dengan mengumpulkan penilaian dari pengguna (dikenal *rating* dalam permasalahan ini) untuk produk/items dalam domain yang diberikan serta pencocokan antara pengguna yang membagi informasi yang dibutuhkan dan sama, serta memiliki kesukaan yang sama [12]. Pada umumnya metode CF yang diterapkan merupakan metode yang berbasis pada ketetanggaan (*neighborhood-based*) yang mana memiliki tiga tahapan sebagai berikut:

- Memberikan bobot ke semua pengguna (data acuan) dengan mempertimbangkan nilai kemiripan dengan pengguna aktif (data yang akan dicari penilaiannya terhadap suatu produk). Pemberian bobot ini dapat dilakukan dengan menerapkan rumusan (2).
- Memilih sebagian data dari pengguna untuk digunakan sebagai pertimbangan prediksi (untuk produk tertentu).
- Melakukan normalisasi serta prediksi penilaian terhadap suatu produk, dengan mempertimbangkan kombinasi bobot pada penilaian pengguna yang terpilih. Prediksi dapat diperoleh dengan mengambil rumusan dari persamaan (6) yang menggunakan kondisi jika pengguna aktif bukan merupakan pengguna baru. Sehingga tahap ini cukup menggunakan persamaan berikut:

$$p_{a,i} = \bar{v}_a + \frac{\sum_{u=1, u \neq a}^n p_{a,u} * (v_{u,i} - \bar{v}_u)}{\sum_{u=1, u \neq a}^n p_{a,u}} \quad (7)$$

Metode CB memberikan rekomendasi dengan melakukan pembelajaran profil secara individu untuk menggambarkan sesuatu, sehingga pemberian rekomendasi cukup dengan mempertimbangkan informasi yang berkaitan dengan produk, tanpa harus mencocokkan minat/kesukaan dari suatu pengguna dengan pengguna lainnya [13]. Pada penelitian ini metode CB terapkan dengan menggunakan langkah-langkah berikut:

- Seluruh deskripsi yang akan digunakan dalam metode CB dilakukan *preprocessing* data terlebih dahulu. Langkah *preprocessing* data diawali dengan penghilangan simbol-simbol tidak berguna dalam deskripsi dengan menggunakan konsep Regular Expression. Selanjutnya *preprocessing* dilanjutkan ke tahap penghilangan kata-kata yang tidak berguna pada konteks klasifikasi, dengan menggunakan acuan *Stopword* bahasa Indonesia.
- Metode CB diterapkan seperti halnya penyelesaian masalah pengkategorian teks. Maka itu, pada penelitian ini digunakan metode untuk klasifikasi teks untuk multi kelas, yaitu *Multinomial Naïve Bayes Text Classification*. Pada dasarnya metode ini menggunakan pendekatan probabilitas untuk menentukan apakah suatu dokumen masuk ke dalam suatu kelas atau tidak, melalui penerapan rumusan berikut:

$$P(c_j|D) = \frac{P(c_j)}{P(D)} \prod_{i=1}^{|D|} P(a_i|c_i) \quad (8)$$

Komponen pada persamaan (8) dapat dipahami sebagai peluang dari suatu dokumen yang diberikan masuk pada suatu kelas yang tergambar pada $P(c_j|D)$, sedangkan $P(c_j)$ adalah peluang kemunculan suatu kelas j , $P(D)$ merupakan peluang kemunculan dokumen yang biasanya dibuat menjadi konstan, $|D|$ merupakan panjang dari dokumen berdasarkan banyak kata yang muncul, serta a_i merupakan kata ke- i dalam suatu dokumen D .

- Berdasarkan perhitungan peluang tersebut, maka keputusan suatu produk diberikan nilai tertentu oleh suatu pengguna adalah melihat deskripsi dari produk (yang dianggap sebagai dokumen) masuk ke dalam suatu penilaian berapa (kelas pada proses klasifikasi). Kelas penilaian yang diambil adalah kelas yang memiliki nilai probabilitas tertinggi dibandingkan kelas lainnya.

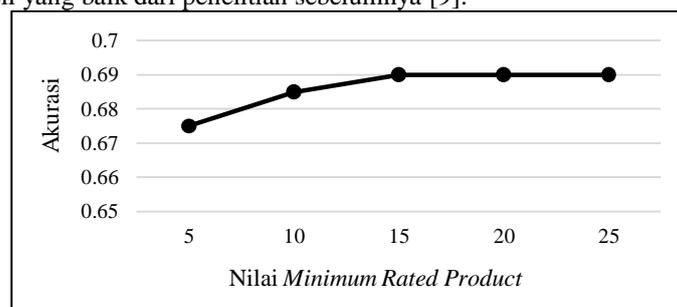
III. HASIL DAN PEMBAHASAN

Proses awal dari implementasi metode CB yaitu *preprocessing* data untuk memperoleh teks yang siap diolah oleh *classifier*. Tahapan ini dianggap penting karena dapat mempengaruhi kualitas pada tahapan klasifikasi dengan memilih kata-kata kunci yang bermakna serta menghilangkan kata-kata yang tidak memiliki kontribusi dalam proses klasifikasi [14]. Terdapat dua teknik yang sering digunakan pada aspek penghilangan kata yang tidak memiliki kontribusi dalam klasifikasi yaitu penghilangan karakter tidak berguna serta penghilangan kata yang tidak berhubungan dengan konteks klasifikasi. Pada penelitian ini, dua teknik tersebut diterapkan dengan tahapan sebagai berikut:

- Pada proses penghilangan karakter yang tidak berguna, diterapkan fungsi *Regular Expression* yang banyak diterapkan pada pemrosesan linguistik dengan melibatkan proses pencocokan pola dari string/kata yang muncul [15]. Melalui penerapan pada Python, penulis menggunakan pola '\W', '\s+[a-zA-Z]\s+' serta '\s+' untuk mendeteksi secara berurutan karakter bukan huruf (*alphanumeric* dan *underscore*), karakter tunggal, serta seluruh penggunaan spasi pada kalimat. Hasil pendeteksian tersebut digunakan untuk proses pengubahan karakter tersebut ke spasi tunggal.
- Selanjutnya teks hasil dari pemrosesan penghilangan karakter tidak berguna dengan fungsi *Regular Expression*, akan dimasukkan ke tahap penghilangan kata-kata yang tidak berhubungan dengan konteks prediksi dengan bantuan fungsi *Stopword* Bahasa Indonesia. Pada penelitian ini digunakan library Sastrawi (berbasis python) untuk pengecekan apakah suatu kata termasuk dalam kata-kata yang tidak berhubungan tersebut. Kata-kata yang masuk kategori tersebut adalah kata yang memiliki frekuensi kemunculan tinggi seperti kata penghubung ('dan', 'atau', 'tetapi', 'akan', dan lainnya).

Tahapan CB pada dasarnya digunakan pengolahan klasifikasi teks dengan multi kelas menggunakan metode *Multinomial Naive Bayes Text Classifier*. Multi kelas yang dimaksud menggambarkan nilai rating yang mungkin bisa diberikan oleh pengguna terhadap produk, yaitu berada pada rentang nilai 0 sampai 5. Selanjutnya penerapan dari metode CF digunakan dua parameter kunci dalam proses perhitungannya, yaitu *minimum rated product* yang akan memberikan pertimbangan yang baik dalam proses perhitungan, serta nilai bobot yang merepresentasikan level kepercayaan tertinggi dari nilai prediksi CB.

Pada penelitian ini digunakan nilai *minimum rated product* yaitu 15, dikarenakan hasil eksperimen yang telah dilakukan terhadap 4 nilai *minimum rated product* (5, 10, 15, 20, 25) diperoleh hasil bahwa parameter dengan nilai 15 memiliki hasil terbaik dari segi akurasi akhir prediksi. Hasil tersebut dapat diamati pada Gambar 2. Sedangkan untuk parameter nilai bobot dari level kepercayaan prediksi CB digunakan nilai 2, karena mempertimbangkan hasil yang baik dari penelitian sebelumnya [9].



Gambar 2. Perbandingan Hasil Akurasi berdasarkan Nilai *Minimum Rated Product*

Pada perhitungan kualitas hasil rekomendasi, digunakan pendekatan lain dari pengujian prediksi yang mana bukan melihat tingkat kecocokan hasil prediksi rating dengan actual rating yang diberikan user. Perhitungan kualitas yang diterapkan menggunakan pendekatan apakah sistem telah baik dalam memprediksi produk bahwa pengguna memang akan menambahkan produk tersebut dalam daftar yang dipertimbangkan. Pendekatan tersebut adalah akurasi prediksi dengan *measuring usage prediction*, yang mana evaluasi dilakukan melalui perbandingan status dipertimbangkannya/digunakannya (data aktual) suatu produk oleh user dengan hasil rekomendasi yang diberikan oleh sistem [16]. Hasil yang akan diperoleh dari perbandingan tersebut dibentuk sebuah tabel dengan 4 kemungkinan hasil, yang mana dapat diamati pada Tabel 1.

Tabel 1. Kombinasi Hasil yang Mungkin dari Proses Rekomendasi

	Produk yang Direkomendasikan	Produk yang Tidak Direkomendasikan
Produk yang Dipertimbangkan	True-Positive (TP)	False-Negative (FN)
Produk yang Tidak Dipertimbangkan	False-Positive (FP)	True-Negative (TN)

Selanjutnya berdasarkan kemungkinan hasil rekomendasi tersebut, dihitunglah nilai dari akurasi, presisi dan recall. Rumusan dan definisi dari 3 penilaian kualitas tersebut adalah sebagai berikut [17]:

- Akurasi merupakan tingkat ketepatan suatu sistem dalam merekomendasikan ataupun tidak merekomendasikan produk secara benar, yang mana diperoleh dengan rumusan (9).

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

- *Recall* atau sensitifitas merupakan proporsi kondisi produk yang direkomendasikan dan sesuai dengan produk yang dipertimbangkan terhadap keseluruhan produk yang dipertimbangkan, yang mana rumusannya sebagai berikut

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

- *Precision* atau tingkat kepercayaan merupakan proporsi kondisi produk yang direkomendasikan dan sesuai dengan produk yang dipertimbangkan terhadap keseluruhan produk yang direkomendasikan, dengan menerapkan rumusan (11) untuk memperolehnya.

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

Berdasarkan penjelasan penentuan parameter dan perhitungan kualitas rekomendasi, serta dengan menerapkan rumusan (12) untuk mengkategorikan nilai rating ke dalam kelompok recommended/used product, maka diperoleh hasil dari kemampuan CBCF dalam pemberian rekomendasi produk dari wedding organizer/vendors serta perbandingan dengan dua algoritma yang umum digunakan yaitu CB dan CF. Hasil tersebut dapat diamati pada Tabel 2.

$$Status\ Rekomendasi\ Produk = \begin{cases} 1 : & \text{Jika nilai rating} \geq 3, \\ 0 : & \text{Selain itu.} \end{cases} \quad (12)$$

Rumusan yang telah dipaparkan sebelumnya, dapat diimplementasikan untuk menghitung kualitas kinerja dari metode yang digunakan. Sebagai contoh, untuk metode CBCF yang digunakan dalam pemberian rekomendasi terhadap pengguna lama diperoleh hasil bahwa 62 data rating yang secara aktual mendapatkan nilai di atas atau sama dengan 3 (tiga), direkomendasikan oleh sistem rekomendasi kepada pengguna (*True Positif*). Selanjutnya, 22 data rating yang secara aktual memperoleh nilai di bawah 3 (tiga), tidak direkomendasikan oleh sistem rekomendasi (*True Negative*). Sedangkan dua kasus kesalahan pemberian rekomendasi, yaitu 13 data rating yang secara aktual mendapatkan nilai di atas atau sama dengan 3 (tiga) namun tidak direkomendasikan oleh sistem (*False Negative*), serta 3 data rating yang secara aktual memperoleh nilai di bawah 3 (tiga) tetapi direkomendasikan oleh sistem (*False Positive*). Maka itu, akan diperoleh nilai kinerja CBCF untuk pemberian rekomendasi kepada pengguna lama, yang dilakukan melalui proses yang dapat diamati pada rumusan (13), (14) dan (15).

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} = \frac{62+22}{62+22+3+13} = \frac{84}{100} = 0.84 \text{ (84\%)} \quad (13)$$

$$Precision = \frac{TP}{TP+FP} = \frac{62}{62+3} = \frac{62}{65} = 0.954 \text{ (95.4\%)} \quad (14)$$

$$Recall = \frac{TP}{TP+FN} = \frac{62}{62+13} = \frac{62}{75} = 0.827 \text{ (82.7\%)} \quad (14)$$

Tabel 2. Hasil Pemberian Rekomendasi dari CBCF dan Perbandingan dengan CB serta CF

Metode	Rekomendasi terhadap Pengguna Lama			Rekomendasi terhadap Pengguna Baru		
	Akurasi	<i>Precision</i>	<i>Recall</i>	Akurasi	<i>Precision</i>	<i>Recall</i>
CBCF	84%	95%	83%	54%	57%	85%
CB	89%	96%	89%	-	-	-
CF	79%	90%	81%	-	-	-

Penulis sadari bahwa hasil yang diperoleh memang tidak terlalu baik, namun dengan mempertimbangkan hasil dari pemberian rekomendasi terhadap pengguna baru, yang mana jika dilakukan dengan metode CF atau CB (biasanya menggunakan profil produk dalam pemberian rekomendasi) maka proses rekomendasi tidak bisa dilakukan. Namun, proses rekomendasi tetap bisa dilakukan dengan CBCF pada kondisi tersebut, dalam penelitian ini untuk kondisi pemberian rekomendasi terhadap pengguna baru, sistem memiliki kemampuan 54% akurasi, 57% *presisi*, serta 85% untuk *recall*.

IV. KESIMPULAN DAN SARAN

Berdasarkan dari hasil dan pembahasan yang telah dipaparkan sebelumnya, maka dapat disimpulkan dalam penelitian ini yaitu:

- CBCF dengan parameter *minimum rated product* yaitu 15 pada penelitian ini memiliki kemampuan yang lebih baik dari nilai lainnya, yang mana CBCF memiliki kemampuan pemberian rekomendasi dengan akurasi sebesar 69%, *precision* sebesar 74%, dan *recall* dengan nilai 84%.
- Kemampuan CBCF memang belum terlalu baik dalam pemberian rekomendasi khususnya kepada pengguna yang telah terdaftar di sistem dan telah memiliki riwayat penilaian. Pada kondisi ini CBCF hanya mampu memperoleh akurasi sebesar 84%, 95% *precision* serta 83% *recall*, yang mana kualitas ini tidak lebih baik dari metode CB.
- Namun, CBCF bisa memberikan rekomendasi pada kondisi pengguna yang baru terdaftar dengan kemampuan akurasi sebesar 54%, *precision* dengan nilai 57% serta 85% *recall* (yang mana jika dengan metode CB ataupun CF kondisi ini tidak bisa dilakukan pemberian rekomendasi).

Berdasarkan hasil, pembahasan serta kesimpulan yang telah diperoleh, banyak pekerjaan berikutnya untuk memperoleh kinerja CBCF yang lebih baik lagi, rencana pengembangan tersebut antara lain:

- Mencoba algoritma lain untuk melakukan klasifikasi teks dengan pertimbangan algoritma yang lebih baik dari *Multinomial Naïve Bayes*, seperti *Support Vector Machine* ataupun varian dari *Artificial Neural Network*.
- Data latih sebaiknya memuat profil pengguna dan produk yang beragam, sehingga membuat proses pembelajaran CBCF menjadi lebih baik, dan lebih bisa memberikan rekomendasi dengan beragam kondisi yang dihadapi.
- Perlu dicoba strategi pembelajaran yang lain dalam melakukan pelatihan CBCF (seperti *K-Fold Cross Validation*), sehingga semua data yang diperoleh pernah diperlakukan sebagai data latih ataupun data uji.

V. UCAPAN TERIMA KASIH

Penelitian ini dibiayai oleh Lembaga Penelitian, Pengabdian kepada Masyarakat, dan Penjaminan Mutu Pendidikan, Institut Teknologi Sumatera (melalui Program Hibah Penelitian ITERA SMART Tahun 2019) sesuai dengan kontrak penelitian B/317/IT9.C1/PT.01.03/2019. Penulis beserta tim juga ingin mengucapkan terima kasih kepada seluruh rekan kerja di Institut Teknologi Sumatera khususnya Program Studi Teknik Informatika atas bantuan keilmuan serta kerjasamanya dalam penelitian ini.

VI. DAFTAR PUSTAKA

- [1] H. K. Soemartono, "Potret Zaman Now Pengguna dan Perilaku Internet Indonesia!," *Buletin APJII*, pp. 1-7, April 2018.
- [2] Anonymous, *The Mobile Economy 2018*, London: The GSM Association, 2018.
- [3] S. Millward, "Indonesia Diproyeksi Lampau 100 Juta Pengguna Smartphone di 2018," *TECHINASIA*, 23 December 2014. [Online]. Available: <https://id.techinasia.com/jumlah-pengguna-smartphone-di-indonesia-2018>. [Accessed 3 March 2019].
- [4] S. M. Maulana, H. Susilo and Riyadi, "Implementasi e-commerce sebagai media penjualan online (studi kasus pada toko pastbrik kota malang)," *Jurnal Administrasi Bisnis*, vol. 29, no. 1, pp. 1-9, 2015.
- [5] S. Kemp, "Digital in 2018: World's Internet Users Pass The 4 Billion Mark," *WeAreSocial*, 30 January 2018. [Online]. Available: <https://wearesocial.com/blog/2018/01/global-digital-report-2018>. [Accessed 6 March 2019].
- [6] D. M. Kristin and Y. Lisanti, "Wedding Organizer Order Management," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 5, no. 2, pp. 839-850, 2014.
- [7] I. Najiyah and Suharyanto, "Sistem Informasi Wedding Planner Berbasis Web," *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, vol. 3, no. 1, pp. 79-86, 2017.
- [8] F. Ricci, L. Rokach and B. Shapira, "Introduction to Recommender System Handbook," in *Recommender System Handbook*, New York, Springer Science+Business Media, 2011, pp. 1-29.
- [9] P. Melville, R. J. Mooney and R. Nagarajan, "Content-Boosted Collaborative Filtering for Improved Recommendations," in *Eighteenth National Conference on Artificial Intelligence*, Edmonton, 2002.
- [10] T. Badriyah, R. Fernando and I. Syarif, "Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori," in *Konferensi Nasional Sistem Informasi (KNSI)*, Pangkal Pinang, 2018.

- [11] L. E. Molina and Fernandez, Recommendation System for Netflix, Amsterdam: Vrije Universiteit Amsterdam, 2018.
- [12] J. L. Herlocker, J. A. Konstan, A. Borchers and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering," in *SIGIR '99: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, 1999.
- [13] R. J. Mooney and L. Roy, "Content-Based Book Recommending Using Learning for Text Categorization," in *SIGIR-99 Workshop on Recommender System: Algorithms and Evaluation*, Berkeley, 1999.
- [14] V. Srividhya and R. Anitha, "Evaluating Preprocessing Techniques in Text Categorization," *International Journal of Computer Science and Application*, vol. 47, no. 11, pp. 49-51, 2010.
- [15] S. Bird, E. Klein and E. Loper, Natural Language Processing with Python, California: O'Reilly Media, Inc., 2009.
- [16] G. Shani and A. Gunawardana, "Evaluating Recommendation Systems," in *Recommender Systems Handbook*, Boston, Springer Science+Business Media, 2011, pp. 257-297.
- [17] D. M. W. Powers, "Evaluation: from Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technology*, vol. 2, no. 1, pp. 37-63, 2011.