

BAB II

TINJAUAN PUSTAKA

2.1 Sistem Rekomendasi

Sistem rekomendasi adalah suatu sistem yang menyarankan informasi yang berguna atau menduga apa yang akan dilakukan pelanggan untuk mencapai tujuannya, misalnya seperti memilih produk tertentu. Sehingga pelanggan memilih produk dapat lebih efektif dalam menentukan produk yang diinginkannya (Kurniawan, 2016).

Sistem rekomendasi telah banyak digunakan oleh hampir sebagian besar area bisnis dimana konsumen perlu membuat suatu keputusan atau rekomendasi pilihan dari informasi yang disediakan. Area pariwisata merupakan salah satu contoh bisnis area yang menerapkan sistem rekomendasi untuk membantu para wisatawan dalam membuat keputusan bagi perjalanan mereka. Internet dan *world wide web* menyediakan banyak informasi dibidang pariwisata karena pariwisata memiliki pengalaman menarik bagi para wisatawan namun sangat sulit untuk menemukan informasi paket wisata yang sesuai dengan keinginan para penggunanya. Maka dibuatlah suatu sistem rekomendasi bagi industri pariwisata atau perjalanan untuk menawarkan serta merekomendasikan paket tempat-tempat wisata di Malang dan sekitarnya kepada para wisatawan yang sesuai dengan keinginan mereka. Sistem rekomendasi didefinisikan sebagai aplikasi pada website *e-commerce* untuk mengusulkan informasi dan menyediakan fasilitas yang diinginkan pengguna dalam membuat suatu keputusan (Ricci, F.,2002). Sistem ini

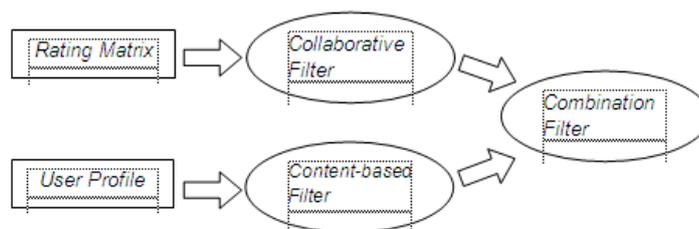
diasumsikan seperti penggambaran kebutuhan dan keinginan pengguna melalui pendekatan metode rekomendasi untuk mencari suatu item dengan menggunakan rating berdasarkan kemiripan dari karakteristik informasi pengguna.

Terdapat beberapa metode yang digunakan untuk mendukung cara kerja sistem rekomendasi dalam menghasilkan sebuah informasi diantaranya seperti *Hybrid Content Based*, *Collaborative Filtering*, *Knowledge Based Recommendation* dan *Association Rule*.

2.1.1 Hybrid Content Based

Penggabungan secara linier (*Hybrid Linear Combination*)

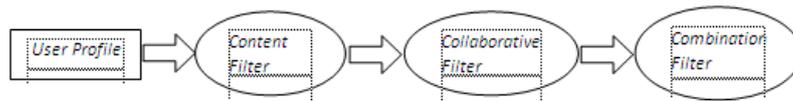
Penggabungan ini menggabungkan hasil prediksi (*rating*) dari metode *contentbased* dan *collaborative*. Penggabungan ini dilakukan dengan cara pemberian ranking atau rating. Penggabungan ini digambarkan pada gambar 2.1.1 berikut :



Gambar 2. 1 *Hybrid Linear Combination*

a. Penggabungan secara sekuensial (*Sequential Combination*)

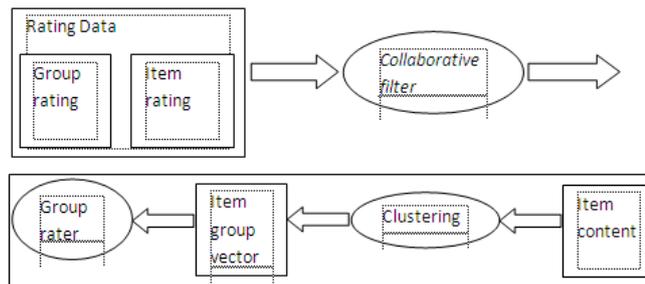
Penggabungan ini adalah melakukan perhitungan pada salah satu metode (misalkan *content-based*) kemudian hasilnya digabungkan dengan metode lainnya (misalkan *collaborative*). Penggabungan ini digambarkan pada gambar 2.1.2 berikut :



Gambar 2. 2 *Hybrid Sequential Combination*

Penggabungan Item-based Clustering Hybrid Method (ICHM)

Penggabungan ini mengintegrasikan informasi item dan rating pengguna untuk menghitung kemiripan item - item. *Item-based clustering Hybrid Method* (ICHM) merupakan sebuah metode yang menerapkan penggabungan *hybrid recommender system* dengan tujuan untuk meningkatkan akurasi prediksi pendekatan *collaborative filtering* dan menangani masalah item baru yang belum *dirating* (*cold-star problem*).



Gambar 2. 3 *Item-based Clustering Hybrid Method*

2.1.2 Collaborative Filtering

Schafer membagi *algoritma collaborative filtering* ke dalam dua kelas yang berbeda, yaitu :

- a) User Based Collaborative Filtering

User based nearest neighbor algorithm menggunakan teknik statistika untuk menemukan sekumpulan pengguna, dikenal sebagai tetangga. Yang memiliki sejarah setuju dengan pengguna yang menjadi sasaran. Setelah sekumpulan tetangga terbentuk sistem menggunakan algoritma

yang berbeda untuk menggabungkan kesukaan *neighbours* untuk menghasilkan predikis atau rekomendasi N-teratas untuk *active user*. (Sarwar dkk, 2001).

b) *Item Base Collaborative Filtering*

Item based collaborative filtering merupakan metode rekomendasi yang didasri atas adanya kesamaan antara pemberian rating terhadap suatu produk dengan produk yang dibeli. Dari tingkat kesamaan produk, kemudian dibagi dengan parameter kebutuhan pelanggan untuk memperoleh nilai kegunaan produk. Produk yang memiliki nilai kegunaan tertinggi adalah yang kemudian dijadikan rekomendasi (Purwanto, 2009). Metode ini muncul sebagai solusi untuk beberapa permasalahan pada *user based collaborative filtering* yaitu pada masalah keterbatasan (*sparsity*) dan skalabilitas serta masalah waktu dan memori. Pada metode *item based collaborative filtering* melakukan similaritas dengan membentuk suatu model similaritas secara *offline* yang secara otomatis akan menghemat waktu dan memori yang digunakan untuk perhitungan pada saat pengguna mengakses halaman situs.

2.1.3 Knowledge Based Recommendation

Knowledge based system merekomendasikan pekerjaan berdasarkan *domain knowledge*, tentang pengaruh fitur dari pekerjaan terhadap kebutuhan dan preferensi pribadi pengguna. Pada kasus ini similarity score pada *knowledge based system* merepresentasikan kesamaan fitur antara pekerjaan yang berbeda. Nilai *similarity* antar

individu bisa didapatkan dengan persamaan. Similarity direpresentasikan dengan *reference matrix* (Rahmawati, Nurjanah, & Rismala, 2018).

2.2 Association Rule

Association rule merupakan salah satu metode yang bertujuan mencari pola yang sering muncul di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa item, sehingga metode ini akan mendukung sistem rekomendasi pengembangan strategi marketing malang melalui penemuan pola antar item dalam transaksi-transaksi yang terjadi di nessashop malang (Wandi, Hendrawan, & Mukhlason, 2012).

Analisis asosiasi atau *Association Rule* adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu.

Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan. Analisis asosiasi juga sering disebut dengan istilah market basket analysis. Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien.

Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, support (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan confidence (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif.

Aturan asosiatif biasanya dinyatakan dalam bentuk :

{roti, mentega} → {susu} (support = 40%, confidence = 50%)

Yang artinya : "50% dari transaksi di database yang memuat item roti dan mentega juga memuat item susu.

Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu." Dapat juga diartikan : "Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini".

Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk support (*minimum support*) dan syarat minimum untuk confidence (*minimum confidence*).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

a. Analisa pola frekuensi tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database. Nilai support sebuah item diperoleh dengan rumus berikut:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}}$$

Sedangkan nilai support dari 2 item diperoleh dari rumus berikut:

$$\text{Support (A} \cap \text{B)} = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Total Transaksi}}$$

b. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiatif $A \rightarrow B$

Nilai confidence dari aturan $A \rightarrow B$ diperoleh dari rumus berikut:

$$\text{Confidence} = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah Transaksi mengandung A}}$$

$P(B | A)$

Sebagai contoh ambil suatu data transaksi yang didapat dari penjualan sayur dengan data transaksi sebagai berikut :

Tabel.2.1 Contoh Data Transaksi

Transaksi	Item yang dibeli
1	Broccoli, Green Peppers, Corn
2	Asparagus, Squash, Corn
3	Corn, Tomatoes, Beans, Squash

Transaksi	Item yang dibeli
4	Green Peppers, Corns, Tomatoes, Beans
5	Beans, Asparagus, Broccoli
6	Squash, Asparagus, Beans, Tomatoes
7	Tomatoes, corn
8	Broccoli, Tomatoes, Green Peppers
9	Squash, Asparagus, Beans
10	Beans, Corn
11	Green Peppers, Broccoli, Beans, Squash
12	Asparagus, Bean, Squash
13	Squash, Corn, Asparagus, Beans
14	Corn, Green Peppers, Tomatoes, Beans, Broccoli

I. Definisi-definisi yang terdapat pada *Association Rule*

1. I adalah himpunan yang tengah dibicarakan.

Contoh: {Asparagus, Beans, ..., Tomatoes}

2. D adalah Himpunan seluruh transaksi yang tengah dibicarakan

Contoh: {Transaksi 1, transaksi 2, ..., transaksi 14}

3. Proper Subset adalah Himpunan Bagian murni

Contoh: Ada suatu himpunan $A = \{a, b, c\}$

Himpunan Bagian dari A adalah

Himpunan Kosong = $\{\}$

Himpunan 1 Unsur = $\{a\}, \{b\}, \{c\}$

Himpunan 2 Unsur = $\{a, b\}, \{a, c\}, \{b, c\}$

Himpunan 3 Unsur = $\{a, b, c\}$

Proper subset nya adalah Himpunan 1 Unsur dan Himpunan 2 Unsur

4. Item set adalah Himpunan item atau item-item di I

Contoh: Ada suatu himpunan $A = \{a, b, c\}$

Item set nya adalah $\{a\}; \{b\}; \{c\}; \{a, b\}; \{a, c\}; \{b, c\}$

5. K- item set adalah Item set yang terdiri dari K buah item yang ada pada I. Intinya K itu adalah jumlah unsur yang terdapat pada suatu Himpunan

Contoh: 3-item set adalah yang bersifat 3 unsur

6. Item set Frekuensi adalah Jumlah transaksi di I yang mengandung jumlah item set tertentu. Intinya jumlah transaksi yang membeli suatu item set.

Contoh: Kita gunakan tabel transaksi penjualan sayur di atas

- frekuensi Item set yang sekaligus membeli Beans dan Broccoli adalah 3
- frekuensi item set yang membeli sekaligus membeli Beans, Squash dan Tomatoes adalah 2

7. Frekuensi Item Set adalah item set yang muncul sekurang-kurangnya “sekian” kali di D. Kata “sekian” biasanya di simbolkan dengan Φ . Φ merupakan batas minimum dalam suatu transaksi

Contoh: Pertama kita tentukan $\Phi = 3$, karena jika tidak di tentukan maka frekuensi item set tidak dapat di hitung.

Jika $\Phi=3$ untuk {Asparagus, Beans} apakah frekuensi Item set?

Jika kita hitung maka jumlah transaksi yang membeli asparagus sekaligus membeli beans adalah 5.

Karena $5 \geq 3$ maka {Asparagus, Beans} merupakan Frekuensi Item set.

8. Fk adalah Himpunan semua frekuensi Item Set yang terdiri dari K item.

II. Langkah-langkah algoritma pada *Association Rule*

1. Tentukan Φ
2. Tentukan semua Frekuensi Item set
3. Untuk setiap Frekuensi Item set lakukan hal sbb:
 - i. Ambil sebuah unsur, namakanlah s
 - ii. Untuk sisanya namakanlah ss-s
 - iii. Masukkan unsur-unsur yang telah di umpamakan ke dalam rule If (ss-s) then s Untuk langkah ke 3 lakukan untuk semua unsur.

2.3 Apriori

Algoritma apriori adalah satu algoritma dasar yang diusulkan oleh Agrawal dan Srikan pada tahun 1994 untuk menemukan frequent itemsets pada aturan asosiasi Boolean. Ide utama pada algoritma apriori adalah : pertama, mencari frequent itemset (himpunan item-item yang memenuhi minimum support.) dari basis data transaksi, kedua – menghilangkan itemset dengan frekuensi yang rendah

berdasarkan level minimum support yang telah ditentukan sebelumnya. Selanjutnya membangun aturan asosiasi dari itemset yang memenuhi nilai minimum confidence dalam basis data (Agrawal & Srikant, 1994). Untuk membentuk kandidat itemset ada dua proses utama yang dilakukan algoritma apriori (Han & Kamber, 2006) :

1. Join Step (Penggabungan) Pada proses ini setiap item dikombinasikan dengan item lainnya sampai tidak terbentuk kombinasi lagi.

2. Prune Step (Pemangkasan) Pada proses ini, hasil dari item yang dikombinasikan tadi kemudian dipangkas dengan menggunakan minimum support yang telah ditentukan oleh user (Dewi Listriani, 2016)

