

BAB II

LANDASAN TEORI

Pada Bab Landasan Teori membahas referensi yang dibutuhkan dalam pembuatan sistem baru yang akan menjadi acuan dalam pengerjaan Tugas Akhir ini dan bermaksud untuk memberikan pengetahuan informasi bagi pembaca. Seperti Informasi mengenai Sistem Rekomendasi dan Data Mining yang membahas keterkaitan teknologi dengan ilmu aspek sastra.

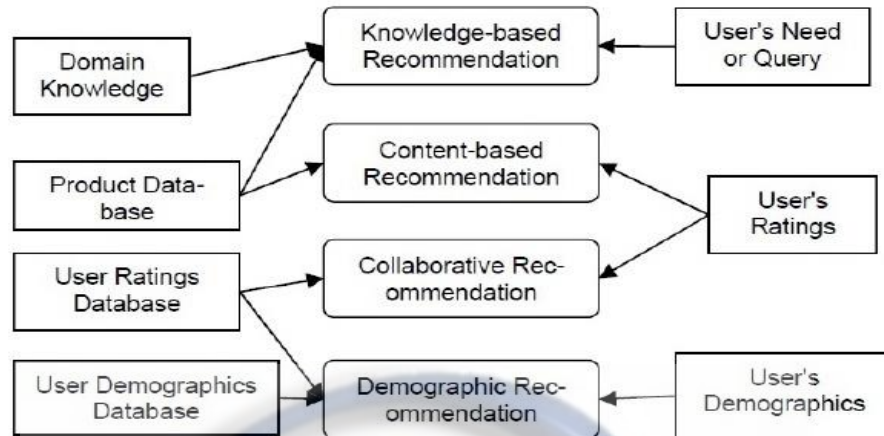
2.1 Sistem Rekomendasi

Sistem rekomendasi merupakan sistem yang bertujuan memperkirakan informasi yang menarik bagi penggunanya dan juga membantu calon konsumen dalam memutuskan barang apa saja yang akan dibelinya^[1] Dalam sistem rekomendasi diperlukan adanya preferensi atau profil pengguna dalam menentukan pilihan dari sekian banyak item yang ada sesuai dengan kebutuhan pengguna. Profil pengguna umumnya didasarkan menarik tidaknya suatu informasi yang dilihat oleh user. Sistem rekomendasi akan menawarkan kemungkinan dari penyaringan informasi personal sehingga hanya informasi yang sesuai dengan kebutuhan dan preferensi pengguna yang akan ditampilkan di sistem dengan menggunakan sebuah teknik atau model rekomendasi.

Ada beberapa metode yang digunakan dalam sebuah sistem rekomendasi. metode atau pendekatan yang dipilih pada sistem rekomendasi bergantung pada permasalahan yang akan diselesaikan, teknik rekomendasi yang berbeda-beda digunakan untuk aplikasi yang berbeda, dasar dari suatu tujuan dan objektif dari sebuah aplikasi. Beberapa metode sistem rekomendasi, yaitu :

1. Content-Based : bekerja dengan cara mencari item lain yang mirip dengan item yang disukai oleh user berdasarkan informasi content atau tekstual dari setiap item,
2. Collaborative Filtering : metode yang digunakan untuk memprediksi kegunaan item berdasarkan penilaian pengguna sebelumnya,
3. Knowledge-Based : merekomendasikan berdasarkan pengetahuan tentang bagaimana item tertentu dapat memenuhi kebutuhan pengguna,

4. Hybrid Based : kombinasi dari beberapa metode yang ada dalam pengembangan sistem rekomendasi.



Gambar 2.1 Diagram teknik rekomendasi dan sumber pengetahuannya menurut Burke, Robin

2.2 Data Mining

Secara sederhana, data mining merupakan ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di database yang besar^[2]. Prosesnya memungkinkan pengguna untuk menganalisis data dari dimensi dan sudut pandang yang berbeda, membuat kategori, dan meringkaskan hubungan yang teridentifikasi.

Gambar 2.2 merupakan komponen utama dalam data mining menurut Han & Kamber (2001) :

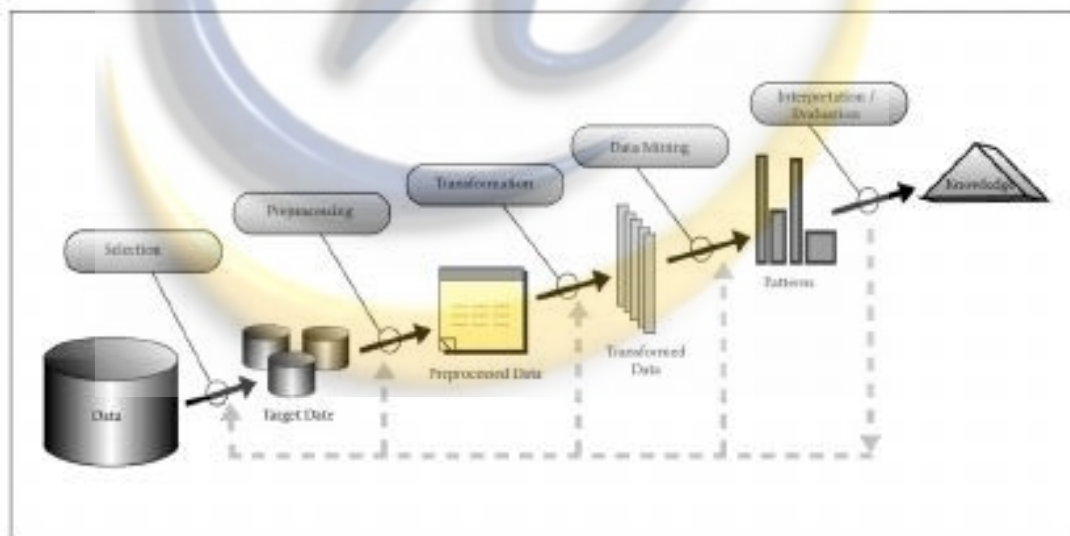


Gambar 2.2 Arsitektur Data Mining

1. Database, data warehouse, internet dan repositori informasi lainnya : merupakan kumpulan data, ataupun jenis informasi lainnya. Pemilahan data atau integrasi data dilakukan untuk pengolahan data.
2. Server database: server database bertanggung jawab dalam pengambilan data yang relevan dengan permintaan pengguna.
3. Knowledge base: digunakan untuk memandu dan mengevaluasi ketertarikan dari pola yang dihasilkan. Pengetahuan tersebut dapat mencakup konsep hirarki yang digunakan untuk mengatur atribut atau nilai atribut ke dalam berbagai tingkat abstraksi.
4. Mesin data mining : hal yang paling penting untuk sistem data mining dan idealnya terdiri dari satu set modul fungsional untuk tugas-tugas seperti karakterisasi, asosiasi dan analisis korelasi, klasifikasi, prediksi, cluster analysis, outlier analysis, dan analisis evolusi
5. Modul evaluasi pola : komponen ini biasanya menggunakan ukuran “ketertarikan” dan interaksinya dengan modul data mining untuk memfokuskan pencarian ke pola yang menarik.
6. Antar Muka (User Interface): modul ini memungkinkan pengguna untuk berinteraksi dengan sistem data mining

. Secara umum, data mining melibatkan enam kelas tugas^[3], yaitu:

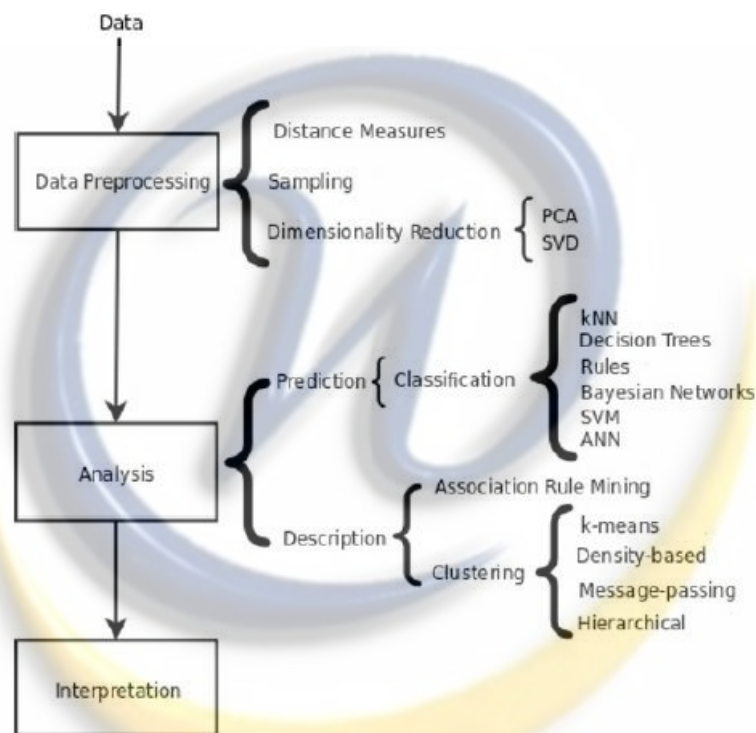
1. Anomaly Detection : identifikasi data yang tidak biasa dari kumpulan data, yang mungkin menarik atau data yang rusak yang membutuhkan investigasi lebih lanjut.
2. Association Rule Learning : mencari hubungan antara variabel data.
3. Clustering : menemukan kelompok – kelompok dan struktur dalam data tanpa menggunakan struktur yang dikenal dalam data.
4. Classification : generalisasi struktur yang dikenal untuk diterapkan ke data baru. Sebagai contoh, program e-mail mungkin akan mencoba untuk mengklasifikasikan e-mail sebagai “sah” atau sebagai “spam”.
5. Regression : mencoba untuk menemukan sebuah fungsi yang memodelkan data dengan kesalahan yang minim.
6. Summarization : memberikan representasi yang lebih padat dari kumpulan data, termasuk visualisasi dan pembuatan laporan



Gambar 2.3 Langkah – langkah dalam menyusun proses KDD^[3]

Sebagian besar teknik yang dipakai dalam sistem rekomendasi merupakan penerapan dari teknik yang dikenal dalam proses data mining . Dalam sistem rekomendasi, data mining digunakan untuk mendeskripsikan kumpulan teknik analisis untuk menyimpulkan aturan rekomendasi atau membangun model

rekomendasi dari kumpulan data yang besar, dan sistem rekomendasi yang menggabungkan teknik tersebut akan menghasilkan rekomendasi berdasarkan pengetahuan yang dipelajari dari tindakan dan atribut pengguna^[4], maka dari itu data mining sering juga disebut knowledge discovery in database (KDD) dimana keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Proses data mining biasanya terdiri dari tiga langkah yang dilakukan secara berurutan^[5], yaitu : data preprocessing, analisis data, dan hasil interpretasi, seperti gambar berikut:



Gambar 2.4 Tiga langkah utama dan beberapa metode dalam Data Mining

Dari Gambar 2.4, maka analisis dalam data mining terdiri dari dua fungsi^[6] yaitu :

1. Fungsi prediksi: memprediksi nilai atribut tertentu berdasarkan nilai atribut yang lain dimana atribut yang diprediksi itu dikenal sebagai variabel yang tergantung pada variabel lain
2. Fungsi deskripsi : memperoleh pola dari kecenderungan korelasi, cluster dan anomali data yang menyimpulkan hubungan dalam data

2.3 Aturan Asosiasi (Association Rule Mining)

Aturan asosiasi adalah teknik untuk menemukan aturan asosiatif antara suatu kombinasi atribut, mencari dan menemukan hubungan antar item yang ada dalam kumpulan data dan bertujuan untuk menemukan informasi item-item yang saling berhubungan dalam bentuk sebuah aturan, maka dari itu, aturan asosiasi termasuk dalam metode analisis market basket. Aturan asosiasi dapat digunakan untuk menemukan “hubungan atau sebab akibat” [7]. Aturan asosiasi ataupun korelasi item- ke-item merupakan salah satu metode yang paling terkenal dalam sistem rekomendasi. Aturan asosiasi telah digunakan dalam berbagai bidang seperti jaringan telekomunikasi, market, manajemen resiko, dan sebagainya.

Market Basket Analysis adalah Analisis terhadap kebiasaan membeli customer dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan customer dalam keranjang belanjanya. Association rule memiliki bentuk LHS→RHS dengan interpretasi bahwa jika setiap item dalam LHS (Left Hand Side) dibeli, maka item dalam RHS (Right Hand Side) juga dibeli [7]. Umumnya terdapat dua ukuran interestingness measure atau nilai ketertarikan yang didapatkan dari hasil pengolahan, yaitu :

1. Nilai pendukung (Support) yaitu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item dari keseluruhan transaksi yang ada dimana ukuran ini menentukan apakah item tersebut layak untuk dicari nilai kepastiannya.
2. Nilai Kepastian (Confidence) yaitu ukuran yang menunjukkan hubungan antar item secara kondisional

Kedua ukuran itulah nantinya yang berfungsi untuk menentukan interesting association rules, untuk dibandingkan dengan batasan yang ditentukan oleh pengguna yang mana batasan tersebut umumnya terdiri dari min_support dan min_confidence [8].

Dalam pencarian aturan asosiasi, diperlukan suatu variabel ukuran yang dapat ditentukan oleh pengguna, untuk mengatur batasan sejauh mana dan sebanyak apa hasil output yang diinginkan oleh pengguna. Dalam asosiasi terdapat istilah

antecedent dan consequent, antecedent untuk mewakili bagian “jika” dan consequent untuk mewakili bagian “maka”. Dimana antecedent dan consequent adalah sekelompok item yang tidak punya hubungan secara bersama ^[6].

2.4 Algoritma Apriori

Algoritma Apriori adalah salah satu teknik yang digunakan untuk menemukan aturan asosiasi. Algoritma apriori adalah algoritma pengambilan data dengan aturan asosiatif untuk menentukan hubungan asosiatif suatu kombinasi item. Algoritma Apriori ini cocok diterapkan bila terdapat beberapa hubungan item yang ingin dianalisa. Untuk penyelesaian masalah asosiasi adalah dengan mengurangi jumlah set item yang dipertimbangkan, lalu pengguna menetapkan nilai minimum support dan nilai min - support. Jadi nilai $(A \cup C) \leq \text{min - support}$ untuk setiap association rule $A \rightarrow C$ atau $(A \rightarrow C) \leq \text{min - support}$. Apriori pertama menghasilkan semua himpunan item yang disebut frequent itemsets.

Terdapat dua proses utama pada algoritma apriori ^[9], yaitu :

1. Penggabungan (Join) Setiap item dikombinasikan dengan item yang lainnya sampai tidak terbentuk kombinasi lagi
2. Pemangkasan (Prune) Hasil kombinasi item akan dipangkas dengan menggunakan minimum support yang telah ditentukan pengguna.

Algoritma apriori bertujuan untuk menemukan frequent itemsets pada kumpulan data. Analisis apriori merupakan suatu proses untuk menemukan semua aturan apriori yang memenuhi syarat minimum untuk nilai pendukung (Support) dan nilai kepastian (Confidence). Nilai pendukung dari sebuah itemset bisa didapatkan dengan rumus :

$$\text{Support (A)} = \frac{\sum \text{Transaksi yang mengandung A}}{\sum \text{Total transaksi yang terjadi}}$$

Sedangkan jika menggunakan dua itemset atau lebih bisa didapatkan dengan rumus :

$$\text{Support (A,B)} = \frac{\sum \text{Transaksi yang mengandung A dan B}}{\sum \text{Total transaksi yang terjadi}}$$

Setelah menentukan nilai support pada itemset berfrekuensi tinggi maka akan dibentuk aturan asosiasi yang menyatakan kuatnya hubungan kombinasi itemset. Untuk menentukan aturan asosiasi dibutuhkan minimal dua itemset, yaitu dengan rumus sebagai berikut :

$$\text{Confidence (A} \rightarrow \text{B)} = \frac{\sum \text{Transaksi yang mengandung A dan B}}{\sum \text{Transaksi yang mengandung A}}$$

Apriori beroperasi dalam dua tahap, tahap pertama semua item dengan dukungan minimum atau item yang sering dihasilkan. Tahap kedua, algoritma menghasilkan aturan dari himpunan semua frekuensi item yang ada.

Frequent itemset merupakan iterasi pada data. Pada iterasi ke-k ditemukan semua himpunan item-item yang mempunyai k item yang disebut k-itemset. Setiap iterasi terdiri dari dua tahap. Pertama, adalah tahap pembangkitan kandidat (candidate generation) dimana himpunan semua frequent (k - 1) -itemset F_{k-1} yang ditemukan pada pass ke-(k - 1) digunakan untuk membangkitkan kandidat itemset C_k. Prosedur pembangkitan kandidat menjamin bahwa C_k adalah superset dari himpunan semua frequent k-itemset. Kemudian data di-scan dalam tahap Penghitungan Support (*Support Counting*). Pada akhir pass C_k diperiksa untuk menentukan kandidat mana yang sering muncul, menghasilkan F_k. Penghitungan support berakhir ketika F_k atau C_{k+1} kosong. Untuk membangkitkan rule akan dibangkitkan lebih dahulu *candidate rule*. *Candidate rule* berisi semua kemungkinan rule yang memiliki *support > minimum support* karena *input candidate rule* adalah *frequent-itemset*. Kemudian *candidate rule* akan *join* dengan tabel F untuk menemukan *support antecedent*. *Confidence rule* dihitung dengan cara membandingkan *support rule* dengan *support antecedent rule*. Hanya *rule* yang mempunyai *confidence > minimum confidence* yang disimpan dalam tabel rule.