

BAB II

TINJAUAN LITERATUR

2.1 Algoritma Apriori

Algoritma apriori menggabungkan jenis aturan asosiasi yang dikenal sebagai afinitas atau analisis keranjang pasar, yang mengidentifikasi hubungan antara beberapa atribut. Salah satu metode untuk menentukan aturan kombinasi item adalah analisis aturan asosiasi. Analisis pola frekuensi tinggi adalah salah satu langkah analisis asosiasi yang telah menarik minat banyak peneliti dalam upaya menciptakan algoritma yang efektif (pola sering). Metrik kemenarikan yang dapat diturunkan dari hasil pemrosesan data menggunakan perhitungan tertentu dapat digunakan untuk mengidentifikasi aturan asosiasi. Beberapa ukuran tersebut adalah :

1. *Support*

Support adalah nilai penunjang atau persentase kombinasi sebuah item dalam database. Rumus *support* :

$$\text{Support}(A \rightarrow B) = P(A \cap B)$$

2. *Confidence*

Nilai kepastian, atau korelasi yang kuat antara item dalam aturan asosiasi, adalah keyakinan. Setelah pola frekuensi kemunculan item diidentifikasi, kepercayaan dapat diselidiki.

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Support}(A \rightarrow B)}{\text{Support}(A)}$$

Atau

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

3. Lift (Interest)

Lift adalah metrik yang mengukur kekuatan hubungan antara hasil independen (preseden-konsekuensi) dan konsekuensinya. Kisaran lift adalah 0 hingga 8. Indikator tidak adanya ketergantungan antara anteseden dan konsekuensi adalah nilai mendekati 1. Nilai yang sangat berbeda dari 1 berarti bahwa anteseden memberi tahu kita sesuatu tentang konsekuensinya.

$$\begin{aligned} \text{Lift}(A \rightarrow B) &= \text{Lift}(B \rightarrow A) \\ \text{Atau} \\ \frac{\text{conf}(A \rightarrow B)}{\text{supp}(B)} &= \frac{\text{conf}(B \rightarrow A)}{\text{supp}(A)} \\ \text{Atau} \\ \text{Lift} &= \frac{P(A \cap B)}{P(A)P(B)} \end{aligned}$$

4. Conviction

Keyakinan adalah metrik yang digunakan untuk mengukur seberapa banyak aturan tersirat. Keyakinan mempertimbangkan dengan cermat bagaimana aturan asosiasi diarahkan. Menurut keyakinan, keyakinan $(A \rightarrow B) \neq$ keyakinan $(B \rightarrow A)$.

$$\begin{aligned} \text{Conviction}(A \rightarrow B) &= \frac{1 - \text{supp}(B)}{1 - \text{conf}(A \rightarrow B)} \\ \text{Atau} \\ \frac{P(A)P(\sim B)}{P(A \cap \sim B)} \end{aligned}$$

5. Leverage

Jumlah item anteseden dan konsekuensi yang dijual secara bersamaan dalam kumpulan data yang lebih besar dari yang diantisipasi diukur dengan metrik yang disebut leverage. Independensi antekeden dan konsekuensinya ditunjukkan dengan nilai 0. Kisaran nilai leverage adalah -0,25 hingga 0,25.

Salah satu metode untuk menentukan aturan asosiasi antar kumpulan item adalah metode aturan asosiasi. Metrik daya tarik berikut dapat diterapkan :

1. *Support*, adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item atau item set dari keseluruhan transaksi.
2. *Confidence*, adalah suatu ukuran yang menunjukkan hubungan antar dua item secara *conditional* (berdasarkan suatu kondisi tertentu).

2.2 Data Mining

Sederhananya, penambangan data adalah proses mengekstraksi informasi baru dari data dalam jumlah besar dengan mencari pola atau aturan tertentu (Davies, 2004). Nama lain untuk data mining adalah seperangkat prosedur untuk menyelidiki nilai tambah dari kumpulan data dalam bentuk pengetahuan yang sebelumnya tidak diketahui secara tangan (Pramudiono, 2007). Nama lain untuk penambangan data adalah penemuan pengetahuan dalam database (KDD). KDD adalah proses pengumpulan dan analisis data historis untuk mencari pola, korelasi, atau keteraturan dalam kumpulan data besar (Santoso, 2007).

Proses mengekstraksi pola yang bermakna dari data dalam jumlah besar— yang dapat disimpan dalam database, gudang data, atau jenis penyimpanan informasi lainnya—dikenal sebagai penambangan data. Sistem basis data, pergudangan data, statistik, pembelajaran mesin, pengambilan informasi, dan komputasi tingkat tinggi adalah beberapa domain ilmiah yang terhubung dengan penambangan data. Selain itu,

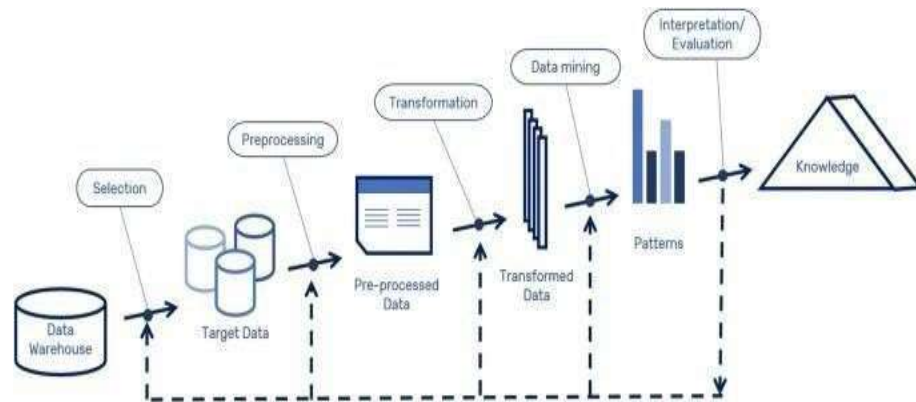
ilmu lain seperti jaringan saraf, pengenalan pola, database gambar, analisis data spasial, dan pemrosesan sinyal membantu penambangan data (Han, 2006). Menemukan pola dalam data adalah proses yang dikenal sebagai penambangan data. Prosedur ini seringkali semi-otomatis atau otomatis. Polanya harus signifikan dan bermanfaat, biasanya menghasilkan keuntungan finansial. Kumpulan data yang signifikan diperlukan (Witten, 2005).

1. Penambangan data adalah proses mengungkap informasi tersembunyi dan pola data yang sebelumnya tidak diketahui.
2. Banyak data biasanya digunakan dalam penambangan data. Big data biasanya digunakan untuk meningkatkan akurasi hasil.
3. Keputusan kritis, terutama yang melibatkan strategi, dapat memperoleh manfaat dari penggunaan penambangan data (Davies, 2004).

Mengingat beberapa wawasan ini, dimungkinkan untuk menarik kesimpulan bahwa penambangan data adalah metode untuk menemukan informasi penting yang tersembunyi dalam kumpulan data yang cukup besar (database) untuk mengungkap pola menarik yang sebelumnya belum ditemukan. Tindakan mengekstraksi sejumlah kecil barang berharga dari sejumlah besar bahan dasar disebut pertambangan. Akibatnya, penambangan data memiliki sejarah panjang dalam domain ilmiah seperti database, pembelajaran mesin, kecerdasan buatan, dan statistik. Literatur penambangan data sering membahas sejumlah teknik, seperti pengelompokan, klasifikasi, penambangan aturan

asosiasi, jaringan saraf, algoritma genetik, dan lain-lain (Pramudiono, 2007).

Pengolahan data dari kumpulan data tidak terstruktur dengan tujuan menemukan pola dalam data adalah definisi dasar dari data mining.



Gambar 2 1 Tahapan Proses Data Mining

1) Data Selection

Sebelum memulai tahap penambangan informasi KDD, data harus dipilih (dipilih) dari kumpulan data operasional. Data dari seleksi yang digunakan untuk proses data mining disimpan dalam file, terpisah dari database operasional.

2) Pre-processing / cleaning

Data yang menjadi fokus KDD perlu dibersihkan sebelum proses penambangan data dapat dilaksanakan. Antara lain, proses pembersihan memerlukan penghapusan data yang berlebihan, memeriksa konsistensi data, dan memperbaiki kesalahan apa pun.

3) Transformation

Proses mengubah data yang dipilih menjadi format yang sesuai untuk penambangan data dikenal sebagai pengkodean. Dalam

KDD, coding adalah proses kreatif yang sangat bergantung pada jenis atau struktur data yang akan dicari di database.

4) Data Mining

Proses penerapan teknik atau metode tertentu pada data yang dipilih untuk menemukan pola atau informasi yang menarik dikenal sebagai data mining. Teknik, metode, dan algoritma penambangan data berbeda secara signifikan berdasarkan tujuan keseluruhan dan proses KDD.

5) Interpretation / evaluation

Pola informasi yang dihasilkan selama proses data mining harus disajikan dengan cara yang mudah dipahami oleh pihak yang berkepentingan. Fase proses KDD ini dikenal sebagai interpretasi. Langkah ini melibatkan penentuan apakah informasi atau pola yang ditemukan bertentangan dengan fakta atau teori yang diketahui.

2.3 Association Rules

Menemukan aturan asosiatif antar item data adalah tujuan dari salah satu tugas penambangan data deskriptif, yang dikenal sebagai aturan asosiasi (Aprianti, Hafizd, & Rizani, 2017). Salah satu teknik yang berusaha mengidentifikasi pola berulang dalam sejumlah besar transaksi adalah aturan asosiasi (Zhang & Zhang, 2002).

Jelas dari definisi di atas bahwa aturan asosiasi adalah teknik untuk mengidentifikasi pola atau aturan lain di antara item tertentu.