**Data Mining**

Definisi dan cara kerja Algoritma Apriori untuk pencarian association rule

Algoritma apriori adalah suatu metode untuk mencari pola hubungan antar satu atau lebih item dalam suatu dataset. Algoritma apriori banyak digunakan pada data transaksi atau biasa disebut *market basket*, misalnya sebuah swalayan memiliki *market basket*, dengan adanya algoritma apriori, pemilik swalayan dapat mengetahui pola pembelian seorang konsumen, jika seorang konsumen membeli item A , B, punya kemungkinan 50% dia akan membeli item C, pola ini sangat signifikan dengan adanya data transaksi selama ini. gue bilang ap? hehe

*Konsep Apriori :
Itemset adalah sekumpulan item item dalam sebuah keranjang (Support)*

*K-itemset adalah itemset yang berisi K item, misalnya beras,telur,minyak adalah 3-itemset (Dinotasikan sebagai K-itemset)*

Frequent support adalah k-itemset yang dimiliki oleh support dimana frequent k-itemset yang dimiliki diatas minimum support atau memenuhi minimum support (dinotasikan sebagai Fi).

Candidat itemset adalah frequent itemset yang dikombinasikan dari k-itemset sebelumnya (dinotasikan sebagi Ci).

Cara kerja apriori :

* Tentukan minimum support
* Iterasi 1 : hitung item-item dari support(transaksi yang memuat seluruh item) dengan men-scan database untuk 1-itemset, setelah 1-itemset didapatkan, dari 1-itemset apakah diatas minimum support, apabila telah memenuhi minimum support, 1-itemset tersebut akan menjadi pola frequent tinggi,
* Iterasi 2 : untuk mendapatkan 2-itemset, harus dilakukan kombinasi dari k-itemset sebelumnya, kemudian scan database lagi untuk hitung item-item yang memuat support. itemset yang memenuhi minimum support akan dipilih sebagai pola frequent tinggi dari kandidat
* Tetapkan nilai k-itemset dari support yang telah memenuhi minimum support dari k-itemset
* lakukan proses untuk iterasi selanjutnya hingga tidak ada lagi k-itemset yang memenuhi minimum support.

Mari kita lihat contoh soal :

Sebuah Supermarket Memiliki data transaksi sebagai berikut



**Gambar 1 : Data Transaksi**

Misal minimum dari nilai support pola frekuensi tinggi adalah 2

* Iterasi 1

untuk 1-itemset hitung dan scan database untuk mendapatkan pola frequent dari support



Gambar 2 : 1-itemset

dapatkan k-itemset dari support yang memenuhi minimum support, kemudian pilih k-itemset sebagai pola frequent tinggi



Gambar 3 : Pola Frequent

* Iterasi 2

pada iterasi sebelumnya pola frequent dari support telah didapatkan dari 1-itemset, untuk 2-itemset, generate k-itemset dari k-itemset iterasi sebelumnya, dengan melakukan kombinasi dari k-itemset tersebut.



Gambar 4: kombinasi dari k-itemset

C2 adalah itemset dari kombinasi k-itemset dari iterasi sebelumnya, setelah didapatkan k-itemset tersebut, hitung masing-masing item frequent dan scan database dan dapatkan frequent item dari support



Gambar 5: 2-itemset

pengembangan algoritma apriori dengan memangkas k-itemset dengan menghitung suppport dari itemset, salah satu itemset yang tidak muncul dalam database {telur,buncis} dari C2, sehingga dipangkas menjadi lebih menghemat memory.

berikut table Pola frequent tinggi diatas minimum support untuk 2-itemset



Gambar 6: Pola frequent tinggi

* Iterasi 3



Gambar 7: Kombinasi dari k-itemset



Gambar 8: 3-itemset dari scan database

kandidat 3-itemset yang telah memenuhi minimum support, itemset tersebut akan menjadi acuan untuk k-itemset selanjutnya



Gambar 9: 3-itemset untuk pola frequent tinggi

* Iterasi 4



Gambar 10: kombinasi 3-itemset untuk k-itemset

scan dabatase untuk mendapatkan itemset dari support, itemset yang memenuhi minimum support dipilih sebagai pola frequent tinggi



Gambar 11: pola frequent tinggi

tidak ada lagi kombinasi yang bisa dibentuk untuk k-itemset berikutnya, proses berhenti, pola frequent tinggi yang ditemukan adalah “roti,mentega,telur,susu”.

Mari kita bentuk association rules yang memenuhi syarat minimum dengan menghitung confidence association rules A->B

Pembentukan Aturan Assosiatif :



Gambar 11 : Aturan Assosiatif

pembentukan aturan assosiatif cukup penting untuk mendapatkan dan menghitung nilai confidence. perlu diketahui algoritma apriori cukup boros dalam penggunaan memory dan paling banyak menghabiskan waktu saat scanning.