

PEMBANGKITAN ITEMSET UNTUK ATURAN ASOSIASI DENGAN ALGORITMA APRIORI DATA MINING

Andreas Chandra

Program Studi Teknik Informatika, STMIK Amikom Yogyakarta
Jl. Ring Road Utara, Condong Catur, Depok, Sleman, DI Yogyakarta 55283
Telp : +62896 7786 2718
E-mail : andreaschaandra@yahoo.com

Abstrak

Data Mining merupakan suatu cara yang digunakan untuk menggali informasi dalam sebuah kumpulan data. Didalam teknik data mining ada banyak metode yang digunakan untuk pengolahan data dimana salah satu metode yang dapat digunakan untuk pengolahan data adalah dengan metode aturan asosiasi. Paper ini membahas tentang pembangkitan item set dimana data-data ini mengacu kepada data transaksi. Pembangkitan itemset ini dilakukan dengan menggunakan aturan asosiasi dengan menggunakan 2 metode pembangkitan yaitu brute force dan $F_{K-1} \times F_1$. Sehingga paper ini lebih memfokuskan kepada analisis cara yang paling cepat untuk pembangkitan itemset, meskipun pembangkitan itemset adalah hal yang masalah yang tidak terlalu penting, namun pembangkitan itemset adalah hal yang berat, pemangkasan kandidat menjadi pekerjaan yang berat karena jumlah itemset yang sangat besar harus diperiksa, dapat dibayangkan apabila terdapat 100 item apabila kita ingin membuat kandidat 2-itemset maka itemset yang dapat dibentuk adalah 1225. Data yang digunakan dalam penelitian ini menggunakan data dari calpoly.edu (<https://wiki.csc.calpoly.edu/datasets/attachment/wiki/apriori/apriori.zip>). Metode brute force adalah metode dimana pembangkitan itemset dilakukan menyeluruh pada item-item yang ada di dataset, sedangkan metode $F_{K-1} \times F_1$ adalah metode dimana itemset dibangkitkan dengan cara menentukan minimum support lalu memfilter itemset mana yang lolos dari nilai yang telah ditetapkan dari minimum support tadi, hal itu akan berlanjut dengan penggabungan 2-itemset dan seterusnya, sehingga proses untuk pembangkitan itemset tidak perlu membuat itemset yang nilai support-nya rendah. Dalam penelitian juga melakukan eksperimen dengan jumlah data. Penulis membandingkan 5 dataset yang berjumlah masing-masing 1000, 2000, 3000, 4000, dan 5000, hasilnya pun berbeda beda, dataset dengan jumlah 1000 membutuhkan waktu 3,619 detik, 2000 dataset membutuhkan waktu 7.223 detik, 3000 dataset membutuhkan 10,896 detik, 4000 dataset membutuhkan 14,380 detik, 5000 dataset membutuhkan 18,087 detik menggunakan metode brute-force untuk mendapatkan 3-itemset, sedangkan menggunakan metode $F_{K-1} \times F_1$ dengan 1000 dataset membutuhkan waktu 0,839 detik, 2000 dataset membutuhkan 2,962 detik, 3000 dataset membutuhkan 5,040 detik, 4000 dataset membutuhkan 8,213, dan 5000 dataset membutuhkan 10,768 detik untuk mendapatkan 3-itemset.

Kata kunci: Data Mining, Apriori, Aturan Asosiasi, Market basket analysis

Halaman ini sengaja dikosongkan