

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis keranjang pasar (juga dikenal sebagai *Association Rule Mining*) merupakan salah satu metode data mining yang berfokus pada menemukan pola pembelian dengan mengekstraksi asosiasi atau kejadian dari data transaksional sebuah toko [1]. Selain itu, analisis keranjang pasar adalah cara yang baik untuk memberikan dukungan keputusan ilmiah pada pasar retail melalui hubungan *mining association* antara barang yang telah dibeli secara bersama-sama [2]. Penemuan dari hubungan ini dapat membantu pedagang untuk mengembangkan strategi penjualan dengan mempertimbangkan barang yang sering dibeli bersama oleh pelanggan [3], hal ini sangat penting karena dapat membantu rekomendasi produk dan promosi produk sehingga strategi pemasaran menjadi lebih tepat. Penentuan pola pembelian barang yang kurang akurat dapat mengakibatkan kebijakan rekomendasi produk dan promosi produk menjadi tidak tepat sasaran

Banyak algoritma yang diusulkan untuk menemukan pola asosiasi [4], dan algoritma pertama yang paling efisien untuk menemukan pola asosiasi adalah *Apriori* [5]. Selain itu banyak juga peneliti-peneliti yang menggunakan algoritma lain, seperti Christidist et al. melakukan penelitian pada *e-commerce* dengan menggunakan *latent topic models* pada data transaksi histori pasar pelanggan sehingga dapat membantu menentukan rekomendasi produk kepada pelanggan menjadi lebih efektif bahkan pada *dataset* dan *itemset* yang besar, Xian Wen et al. menentukan *association rule* dengan menggunakan *matrix database*.

Chien-Hua Wang et al [6]. mengatakan bahwa dalam algoritma *association rule* konvensional, pemindaian database membutuhkan waktu yang sangat besar terutama ketika seseorang menggunakan Algoritma *Apriori*, yang sering mempengaruhi efisiensi dalam data mining. Untuk mengatasi kelemahan tersebut, Han et al. mengusulkan metode data mining, yang disebut *Frequent Pattern-Growth (FP-*

Growth), yang tidak perlu untuk menghasilkan *generate candidate item sets* dan dianggap lebih efisien. *FP-Growth* dibangun dengan membaca kumpulan data satu transaksi pada waktu dan pemetaan setiap transaksi ke dalam bagian *Frequent Pattern-Tree (FP-Tree)*.

Untuk menganalisis keranjang pasar, pendekatan yang biasa digunakan adalah aturan asosiasi. Tetapi ada sejumlah masalah teknis yang berhubungan dengan teknik rekomendasi yang paling umum. Aturan asosiasi cenderung mengabaikan *itemset* besar, dan rekomendasi *item* kurang tepat karena informasi tentang produk ritel tidak tersedia [7], sehingga untuk data yang besar hasilnya menjadi kurang akurat. Untuk mengatasi masalah tersebut, atribut yang ada di *cluster* untuk membentuk kelompok atribut yang sama dan kemudian menentukan pola asosiasi pada masing-masing kelompok [5], sehingga dapat mempermudah proses mencari rekomendasi produk.

Metode untuk melakukan *clustering* dapat dikategorikan menjadi empat metode, yaitu: *partitioning*, *hierarchical*, *grid-based* and *model-based*. *Clustering* berbasis *partitioning* menghasilkan partisi dari data sehingga objek dalam *cluster* lebih mirip satu sama lain daripada objek yang ada dalam *cluster* lain. *K-Means* dan *K-Medoids* adalah contoh dari metode *partitioning*. Algoritma *K-Means* sensitif terhadap *outlier* karena objek dengan nilai yang sangat besar dapat secara substansial mendistorsi distribusi data. Untuk mengambil nilai rata-rata dari objek dalam sebuah *cluster* sebagai titik acuan, *medoid* dapat digunakan, yang merupakan objek dalam sebuah *cluster* yang paling terpusat. Strategi dasar dari algoritma *clustering K-Medoids* adalah untuk menemukan k *cluster* dalam n objek dengan pertama kali secara *arbitrarily* menemukan wakil dari objek (*medoid*) untuk tiap-tiap *cluster* [8].

Penelitian ini, akan menggunakan algoritma *K-Medoids* untuk *clustering* pada data penjualan dan menerapkan algoritma *FP-Growth* untuk pendekatan asosiasi pada setiap *cluster*. Sehingga rekomendasi produk kepada pelanggan menjadi lebih akurat dikarenakan *dataset* yang akan diasosiasi menjadi lebih kecil.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas, masalah pada penelitian ini adalah *FP-Growth* banyak digunakan untuk rekomendasi produk, tetapi untuk jumlah data yang besar, aturan asosiasi menjadi kurang akurat. Untuk mengatasi permasalahan jumlah data yang besar, dapat dilakukan proses *clustering* data dengan menggunakan *K-Medoids* sebelum menentukan pola asosiasi.

Pertanyaan penelitian pada penelitian ini adalah berapa *rule* valid yang dihasilkan melalui *FP-Growth* untuk menentukan rekomendasi produk pada dataset yang besar apabila diterapkan *K-Medoids* pada proses *clustering* data?

1.3 Tujuan Penelitian

Tujuan penelitian ini menerapkan *K-Medoids* untuk *clustering* pada *FP-Growth* dalam menghasilkan *rule* valid untuk menentukan rekomendasi produk pada dataset yang besar.

1.4 Manfaat Penelitian

1.4.1 Manfaat bagi Mahasiswa

Merupakan sarana latihan praktis bagi mahasiswa dalam menerapkan ilmu yang telah diperoleh selama kuliah.

1.4.2 Manfaat bagi Organisasi Bisnis

Hasil penelitian ini diharapkan dapat digunakan untuk pengambilan keputusan dalam membantu menentukan kebijakan, khususnya dalam bidang strategi pemasaran.

1.4.3 Manfaat bagi IPTEKS

Memberikan sumbangan bagi pengembangan teori yang berkaitan dengan proses pelatihan dari *FP-Growth* dan penerapan penggunaan *K-Medoids* sebagai metode *clustering* untuk menentukan rekomendasi produk.

1.5 Kontribusi Penelitian

Integrasi *K-Medoids* dan *FP-Growth* berhasil menghasilkan *rule* untuk rekomendasi promosi produk dengan tingkat kevalidan yang tinggi pada dataset besar.