

ANALISIS KERANJANG PASAR MENGGUNAKAN K-MEDOIDS DAN FP-GROWTH

Wiwit Agus Triyanto¹, Vincent Suhartono², H. Himawan³

^{1,2,3} Pascasarjana Teknik Informatika Universitas Dian Nuswantoro

aguswiwit@yahoo.com
vincentsuhartono@yahoo.com
himawan83@yahoo.com

Abstrak: Analisis keranjang pasar (juga disebut dengan penambangan aturan hubungan) adalah salah satu metode penambangan data yang memfokuskan pada penemuan pola pembelian dengan mengekstrak hubungan atau kejadian dari data transaksi pasar. Mendapatkan pola pembelian sangat penting karena dapat membantu dalam menyusun strategi rekomendasi dan promosi produk. Ada banyak algoritma yang dapat digunakan untuk menemukan pola hubungan, seperti *Apriori* dan *FP-Growth*. Namun ada beberapa kendala teknis yang berhubungan dengan teknik rekomendasi yang biasa digunakan, aturan hubungan sering kali mengabaikan kumpulan item yang banyak. Untuk mengatasi hal ini, atribut yang ada dikelompokkan berdasarkan atribut yang sama dan kemudian ditentukan hubungan antar pola dalam setiap grup. Penelitian ini akan menggunakan algoritma *K-Medoid* untuk pengklusteran data penjualan dan menerapkan algoritma *FP-Growth* untuk mendekati hubungan tiap kluster. Sehingga rekomendasi produk kepada konsumen dapat lebih akurat karena kelompok data yang akan dihubungkan menjadi lebih kecil. Nilai minimum yang mendukung eksperimen yaitu 10%-100% dan nilai minimum kepercayaan adalah 10%-100%. Dari pengukuran menggunakan rasio dukungan, kepercayaan, dan kenaikan ditemukan bahwa terdapat banyak aturan yang salah pada kluster kelima.

Kata kunci: Analisis Keranjang Pasar, Rekomendasi Produk, Promosi Produk, *Association Rule Mining*, *K-Medoids*, *FP-Growth*

Abstract: Market basket analysis (also known as Association Rule Mining) is one method of data mining is focusing on discovering the buying patterns by extracting associations or events from a transactional data store. Purchase pattern discovery is very important because it can help recommend products and promotional products to be more precise marketing strategy. To analyze market basket approach is commonly used association rules. Many algorithms to find patterns of associations, such as *Apriori* and *FP-Growth*. But there are a number of technical problems associated with the most common recommendation techniques, association rules tend to ignore the large itemset. To overcome these problems, existing attributes clustered to form groups of the same attributes and then determine the association patterns in each group. This study will use *K-Medoids* algorithm for clustering on sales data and apply the *FP-Growth* algorithm to approach the association in each cluster. So that the product recommendations to customers to be more accurate because the dataset that will be associated to be smaller. To the experimentally determined value of Minimum Support is 10% - 100% and Confidence Minimum value 10% - 100%. From the measurement results using Support, Confidence

and Lift Ratio is found that a high number of invalid rule in the fifth cluster.

Keywords: Market Basket Analysis, Recommendation Product, Promotion Product, Association Rule Mining, *K-Medoids*, *FP-Growth*.

I. PENDAHULUAN

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, 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.

II. TINJAUAN PUSTAKA

Berikut penjabaran dari penelitian-penelitian yang terkait :

- a. Judul: *Exploring Customer Preferences with Probabilistic Topics Models* [7]

Christidist et al menganalisis keranjang pasar pada *e-commerce* untuk menentukan rekomendasi produk kepada pelanggan. Pendekatan yang umum digunakan adalah aturan asosiasi, tetapi ada sejumlah masalah, yaitu: aturan asosiasi cenderung mengabaikan *itemset* yang besar dan rekomendasi produk kurang tepat karena tidak ada informasi tentang produk ritel. Mereka mencoba menerapkan metode *latent topic*, karena metode ini lebih memperhatikan *itemset* dan juga histori penjualan produk.

Metode yang digunakan adalah 5 topic model dan asosiasi menggunakan algoritma *FP-Growth*. Topik-topik model yang digunakan, yaitu:

- 1) *Latent Baskets – Gibbs Sampler* untuk memprediksi perilaku pelanggan berdasarkan dari produk-produk yang ada di keranjang pasar pelanggan.
- 2) *Latent Baskets Theasurus* untuk mengkalkulasi produk yang mirip dengan histori keranjang pasar pelanggan tersebut sampai produk yang berbeda. Kemudian membandingkan produk relevan dengan produk yang benar-benar dipilih oleh pelanggan tersebut.
- 3) *Latent Basket with Co-occurrence Boosting*. Hampir sama dengan *Theasurus*, yaitu mengkalkulasi jumlah produk yang mirip dengan histori keranjang pasar pelanggan.
- 4) *Latent Users Theasurus* untuk memprediksi produk yang akan dibeli yaituberdasarkan

produk-produk yang ada di keranjang pasar pelanggan.

- 5) *Latent Baskets combined with Latent Users*.

Hasil evaluasi menyimpulkan bahwa metode *Latent Baskets – Gibbs Sampler* adalah yang paling efektif untuk rekomendasi produk kepada pelanggan dibandingkan dengan metode lainnya.

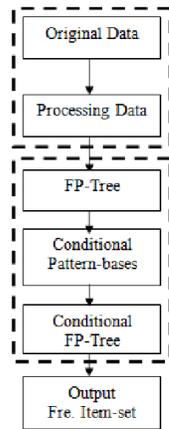
- b. Judul: *Improved Algorithms Research for Association Rule Based on Matrix* [10]

Meskipun algoritma *Apriori* menggunakan *cut-technology* ketika *generate candidate itemset*, hal tersebut harus dilakukan setiap saat dengan cara melakukan scan database pada setiap transaksi. Kecepatan *scanning* akan menjadi lambat untuk jumlah data yang besar. Peningkatan algoritma *Apriori* berbasis *algoritma matrix*, memiliki ide dasar dengan mengubah kejadian dalam database ke dalam *database matriks* sehingga untuk mendapatkan *item matriks* ditetapkan *itemset* yang maksimal. Ketika menemukan *frequent k-item set* dari *frequent k-item set*, hanya set matriks ditemukan. Jadi hanya data yang sesuai dihitung untuk mendapatkan *frequent k-item set* yang ditetapkan. Oleh karena itu peningkatan algoritma *Apriori* membutuhkan waktu komputasi sangat cepat. Hasil percobaan membuktikan efisiensi peningkatan algoritma *Apriori*.

- c. Judul: *FP-Growth Algorithm for Application in Research of Market Basket Analysis* [9]

Association Rules adalah konten penelitian yang penting dalam data mining. Tetapi produksi *frequent set* adalah langkah pertama sebelum produksi aturan asosiasi. Pada kenyataannya banyak orang umumnya menggunakan algoritma seperti *Apriori*.

Apriori memiliki kekurangan yang sangat besar, kita harus mendapatkan *frequent sets* yang diproduksi *candidate frequent sets* tanpa henti. Namun, *cost* dari *candidate frequent sets* sangat luar biasa. Kelebihan algoritma *FP-Growth* adalah bahwa hal itu dapat menghemat waktu dan ruang penyimpanan dan penggunaan sarana partisi untuk menghindari database skala besar. Untuk algoritma *FP-Growth*, biasanya membangun *frequent sets* melalui *FP-Tree*.



Gambar.10. Metode yang diusulkan Yongmei Liu [11]

Ketika database begitu besar *FP-Tree* masih belum dibangun dalam memori. setelah database dipecah menjadi subset data yang dengan *frequent 1-set*, maka penggalan data akan sulit untuk dilaksanakan dengan lancar. Ide meningkatkan adalah bahwa bagian data yang dengan *frequent 1-set* dipecah menjadi bagian data dengan *frequent 2-set*. Jika *FP-Tree* masih belum dibangun dalam memori setelah dipecah, subset data tidak akan rusak sampai *FP-Tree* dibangun.

2.1 Landasan Teori

a. Analisis Keranjang Pasar

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].

Analisis keranjang pasar bermula dari transaksi-transaksi yang berisi satu atau lebih barang/item, dan beberapa informasi sementara dari transaksi tersebut. Untuk melakukan analisis keranjang pasar, berikut langkah-langkah:

- 1) Tentukan nilai *Minimum Support* yang diinginkan. *Minimum Support* merupakan ambang batas minimum jumlah *itemset* yang diperbolehkan, jika jumlah *itemnya* di bawah ambang batas maka *item* tersebut akan dieliminasi.
- 2) Menetapkan *frequent itemset* (kumpulan *item* yang muncul secara bersamaan), dengan cara mengambil *itemset* yang memiliki frekuensi *itemset* minimal sebesar *Minimum Support* sebelumnya.
- 3) Dari semua *frequent itemset*, hasilkan aturan asosiasi yang memenuhi nilai *Minimum Support*.

Analisis keranjang pasar didasarkan pada tiga matrik: *Support*, *Confidence* dan *Lift*. Ketiga matrik tersebut berasal dari catatan transaksi untuk bisnis [12].

1) Support

Matrik pertama ditetapkan untuk analisis keranjang pasar adalah *Support*, yang merupakan probabilitas dari asosiasi (probabilitas dari dua *item* yang dibeli bersamaan). *Support* dihasilkan dari berapa kali jumlah *item* A dan B terjadi bersamaan dalam transaksi yang sama dibagi dengan jumlah total dari transaksi tersebut. *Support* dapat dirumuskan sebagai berikut:

$$\begin{aligned}
 \text{Support} &= P(A \cap B) \\
 &= \frac{\text{jumlah transaksi yang memuat A dan B}}{\text{total jumlah transaksi}}
 \end{aligned}$$

2) *Confidence*

Confidence dihasilkan dari seberapa kuat hubungan produk yang sudah dibeli. *Confidence* dapat dirumuskan sebagai berikut:

$$Confidence = P(B/A) = \frac{Support(A \cap B)}{P(A)}$$

3) *Lift*

Lift Ratio mengukur seberapa penting *rule* yang telah terbentuk berdasarkan nilai *support* dan *confidence*. *Lift Ratio* merupakan nilai yang menunjukkan kevalidan proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B. *Lift Ratio* dapat dirumuskan sebagai berikut:

$$Lift\ Ratio = \frac{Support(A \cap B)}{Support(A) * Support(B)}$$

b. Clustering

Mengelompokkan *record*, pengamatan dan membentuk kelas obyek-obyek yang memiliki kemiripan. Tujuan dari algoritma *cluster* adalah dengan memecahkan setiap data dalam *dataset* menjadi kelompok-kelompok yang homogen. Kelompok data ini biasanya disebut sebagai *cluster*. Setiap *cluster* yang terbentuk akan terdiri dari data yang sejenis dan berbeda dengan data pada *cluster* lainnya [14].

Pengelompokkan ini sama dengan cara kerja otak manusia, dimana ilmu pengetahuan dikelompokkan dalam setiap bidangnya. Dengan adanya pengelompokkan, data yang dapat diolah dengan lebih spesifik sesuai dengan tujuan penelitian. Pemecahan data ke dalam *cluster* data juga diterapkan pada tahap pengolahan awal data dalam proses data mining, sehingga dapat diterapkan metode data mining data mining ke dalam setiap *cluster* data. Proses *clustering* juga dapat mengurangi jumlah ataupun dimensi data yang diolah.

c. Asosiasi

Analisis asosiasi atau *association Rule mining* adalah teknik data *mining* untuk menemukan aturan asosiatif antara suatu kombinasi *item*. Contoh dari aturan *assosiatif* dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu. Karena analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan, analisis asosiasi juga sering disebut dengan istilah analisis keranjang pasar (*market basket analysis*) [15].

Contohnya yaitu pada sebuah supermarket ditemukan 1000 pelanggan yang berbelanja pada hari kamis malam [13], dimana:

- 200 membeli popok
- dari 200 yang membeli popok, 50 dari mereka juga membeli bir.

Sehingga pola asosiasinya yaitu menjadi “Jika beli popok, maka beli bir” dengan perbandingan $200/1000 = 20\%$ dan $50/200 = 25\%$.

d. Algoritma FP-Growth

Algoritma yang biasa dipakai dalam mencari *frequent itemset* antara lain algoritma *Apriori* dan algoritma *FP-Growth*. Pada penelitian ini akan dibahas bagaimana pencarian *frequent itemset* menggunakan algoritma *FP-Growth*. *FP-Tree* (*Frequent Pattern Tree*) digunakan bersamaan dengan algoritma *FP-Growth* untuk menentukan *frequent itemset* (data yang paling sering muncul) dari sebuah *dataset*.

Algoritma *Apriori* memerlukan langkah *candidate generation*, yaitu dengan melakukan

scanning dataset secara berulang-ulang untuk menentukan *frequent itemset*. Algoritma *FP-Growth* adalah salah satu cara alternatif untuk menemukan himpunan data yang paling sering muncul (*frequent itemset*) tanpa menggunakan generasi kandidat [16].

Untuk menentukan *frequent itemset* pada data transaksi tersebut, dapat dilakukan langkah-langkah berikut ini:

- 1) Menentukan *Minimum Support*
- 2) Menentukan *Header Frequent Itemset*
- 3) Membuat *FP-Tree*
- 4) Membuat *Conditional Pattern* berdasarkan *FP-Tree*
- 5) Menentukan *Frequent Item-set*

e. Algoritma *K-Medoids*

Untuk melakukan *clustering* dengan metode partisi dapat menggunakan *K-Means* dan *K-Medoids*. *K-Means* merupakan suatu algoritma pengclusteran yang cukup sederhana yang mempartisi *dataset* kedalam beberapa *cluster* k . Algoritmanya cukup mudah untuk diimplementasi dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan [16].

Algoritma *K-Medoids* hadir untuk mengatasi kelemahan Algoritma *K-Means* yang sensitif terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data [17].

Menurut Han dan Kamber, algoritma *K-Medoids* adalah sebagai berikut [17].

- 1) Secara acak pilih k objek pada sekumpulan n objek sebagai *medoid*.
- 2) Ulangi:
- 3) Tempatkan objek *non-medoid* ke dalam *cluster* yang paling dekat dengan *medoid*.
- 4) Secara acak pilih Orandom: sebuah objek *non-medoid*.

- 5) Hitung total *cost*, S , dari pertukaran *medoid* oj dengan Orandom.
- 6) Jika $S < 0$ maka tukar oj dengan Orandom untuk membentuk sekumpulan k objek baru sebagai *medoid*.
- 7) Hingga tidak ada perubahan.

2.2 Kerangka Pemikiran

Berdasarkan pandangan diatas maka kerangka pemikiran yang dihasilkan (lihat Appendiks 1).

Pada penelitian ini, digunakan data transaksi penjualan dari yang bersifat *public*. Metode yang diusulkan adalah menggunakan algoritma *K-Medoids* untuk *clustering* pada data penjualan dan menerapkan algoritma *FP-Growth* untuk pendekatan asosiasi pada setiap *cluster*.

Untuk meningkatkan akurasi menggunakan aturan asosiasi dengan *FP-Growth* pada *dataset* yang besar, maka *dataset* *diclustering* dahulu menjadi 5 *cluster*. Tujuannya untuk mengecilkan *dataset* dalam proses asosiasi.

Masing-masing dari *cluster* yang terbentuk akan dilakukan proses asosiasi menggunakan algoritma *FP-Growth* untuk menentukan rekomendasi produk kepada pelanggan. Hasil dari proses asosiasi ini diukur menggunakan *Support*, *Confidence*, *Lift Ratio*. Akurasi dari penerapan algoritma *FP-Growth* yang *diclustering* dahulu menggunakan *K-Medoids* akan dibandingkan dengan akurasi dari penerapan algoritma *FP-Growth* saja.

III. METODE PENELITIAN

3.1. Desain Penelitian

Penelitian ini menggunakan jenis penelitian eksperimen, dengan tahapan penelitian sebagai berikut:

1) Pengumpulan Data

Pengumpulan data merupakan langkah awal pada suatu penelitian. Data yang digunakan pada penelitian ini adalah data penjualan supermarket.

2) Pengolahan Data Awal

Pengolahan awal (*Preprocessing*) merupakan tahap untuk mempersiapkan data yang telah diperoleh dari tahap pengumpulan data, yang akan digunakan pada tahap selanjutnya.

3) Desain Eksperimen

Tahapan ini akan membahas desain eksperimen yang digunakan pada penelitian.

4) Eksperimen dan Pengujian

Tahapan ini akan membahas tahapan penelitian dan teknik pengujian yang akan digunakan.

5) Evaluasi Penelitian

Tahapan ini akan membahas hasil evaluasi dari eksperimen yang telah digunakan.

3.2. Pengumpulan Data

Teknik pengumpulan data dilakukan secara kuantitatif, dengan mengambil *dataset* analisis keranjang pasar. Dalam *dataset* ini mendefinisikan transaksi penjualan pada sebuah supermarket. Terdapat 304 atribut yang 303 merupakan *item*/barang dan salah satunya merupakan ID transaksi penjualan. Terdiri dari 1361 record transaksi.

Tabel 6. Contoh *Dataset*

| Bas- ket ID | Lemo- ns | Stan- dard coffee | Frozen Chicken Wings | 98pct. Fat Free Ham- burger | Sugar Coo- kies | Onions |
|----------------|-------------|-------------------------|----------------------------|---|-----------------------|--------|
| C11867 | false | False | true | true | false | false |
| C5096 | false | False | false | true | false | true |
| C4295 | false | False | false | false | false | true |
| C2837 | true | False | false | false | true | false |
| C2693 | true | False | true | false | false | true |
| C3497 | false | True | false | true | false | false |

Keterangan:

- 1) **Basket ID** merupakan ID Transaksi penjualan barang.
- 2) **Lemons, Standard coffee, Frozen Chicken Wings, 98pct. Fat Free Hamburger, Sugar Cookies, dan Onions** merupakan barang-barang yang dijual di Supermarket.
- 3) **False** menandakan bahwa barang tersebut tidak dibeli oleh pembeli.
- 4) **True** menandakan bahwa barang tersebut dibeli oleh pembeli.

3.3. Pengolahan Data Awal

Data yang sudah dikumpulkan akan diolah dengan beberapa tahap sehingga menjadi sebuah dataset yang akan diimplementasikan terhadap metode-metode data mining. Semua *record* transaksi yang ada pada dataset penjualan digunakan dalam tahap pengolahan data. Tahapan pengolahan data yang dilakukan yaitu:

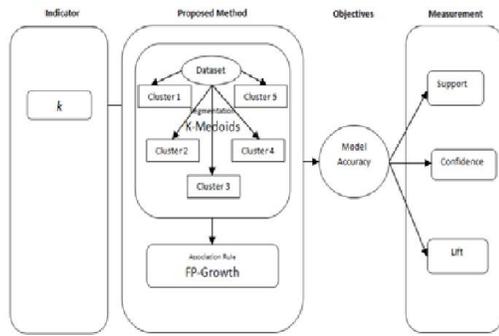
- 1) Pengubahan status barang yang semula *true / false* menjadi 1 / 0, yang akan digunakan untuk proses asosiasi.
- 2) Penghitungan jumlah barang yang dibeli maupun yang tidak dibeli dilakukan dengan menambah 2 atribut di dalam dataset

| Basket ID | Jumlah Barang yang dibeli | Jumlah Barang yang tak dibeli | Hair Conditioner | Lemons | Standard coffee |
|-----------|---------------------------|-------------------------------|------------------|--------|-----------------|
| C11867 | 1 | 302 | 0 | 0 | 0 |
| C5096 | 2 | 301 | 0 | 0 | 0 |
| C4295 | 1 | 302 | 0 | 0 | 0 |
| C2837 | 1 | 302 | 0 | 0 | 0 |
| C2693 | 1 | 302 | 0 | 0 | 0 |
| C3497 | 1 | 302 | 0 | 0 | 0 |
| C2696 | 1 | 302 | 0 | 0 | 0 |
| C1885 | 17 | 285 | 0 | 0 | 0 |
| C9524 | 1 | 302 | 0 | 0 | 0 |

Gambar 2. Menghitung Jumlah Beli / Tidak beli

3.4. Desain Eksperimen

Eksperimen pada penelitian ini adalah penggunaan *K-Medoids* untuk *clustering* data dan *FP-Growth* untuk proses asosiasi.



Gambar 3. Desain Eksperimen

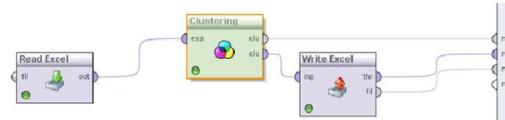
Eksperimen menggunakan algoritma *K-Medoids* untuk *clustering* pada dataset dan menerapkan algoritma *FP-Growth* untuk pendekatan asosiasi pada setiap *cluster*. Sebelum data yang sudah dikumpulkan diolah untuk menemukan pola asosiasi dengan menggunakan *FP-Growth*, terlebih dahulu dilakukan proses *clustering* data dengan menggunakan *K-Medoids*. *Clustering* data dilakukan dengan cara membagi data yang ada menjadi 5 *cluster* (*k*) berdasarkan jumlah barang yang dibeli maupun yang tidak dibeli. Tujuan dari *clustering* adalah agar data yang di asosiasi menjadi lebih kecil sehingga pola yang dihasilkan dapat lebih akurat.

Masing-masing dari *cluster* yang terbentuk akan dilakukan proses asosiasi menggunakan algoritma *FP-Growth* untuk menentukan rekomendasi produk kepada pelanggan. Hasil dari proses asosiasi ini diukur menggunakan *Support*, *Confidence*, *Lift Ratio*. Akurasi dari penerapan algoritma *FP-Growth* yang di *clustering* dahulu menggunakan *K-Medoids* akan dibandingkan dengan akurasi dari penerapan algoritma *FP-Growth* saja.

3.5. Eksperimen dan Pengujian dengan RapidMiner

Dalam melakukan penelitian ini diperlukan eksperimen dan proses pengujian model yang

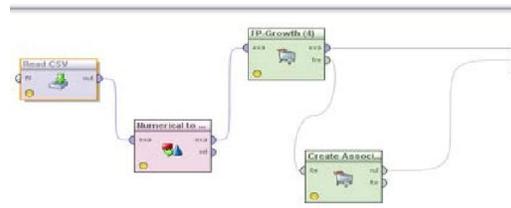
diusulkan. Proses eksperimen dan pengujian menggunakan bagian dari dataset yang ada. Sebelum eksperimen dilakukan, dataset yang ada harus sudah di *clustering* menjadi beberapa *cluster*. Setiap *cluster* yang dihasilkan dari proses *clustering* akan dilakukan asosiasi dengan *FP-Growth* untuk menghasilkan rekomendasi produk.



Gambar 4. Proses Clustering Data

| Row No. | Basket ID | cluster | Jumlah Barang yang dibeli | Jumlah Barang yang tak dibeli |
|---------|-----------|-----------|---------------------------|-------------------------------|
| 1 | C11857 | cluster_0 | 1 | 302 |
| 2 | C5096 | cluster_0 | 2 | 301 |
| 3 | C4295 | cluster_0 | 1 | 302 |
| 4 | C2837 | cluster_0 | 1 | 302 |
| 5 | C2693 | cluster_0 | 1 | 302 |
| 6 | C3497 | cluster_0 | 1 | 302 |
| 7 | C2696 | cluster_0 | 1 | 302 |
| 8 | C1895 | cluster_3 | 17 | 286 |
| 9 | C9524 | cluster_0 | 1 | 302 |
| 10 | C9527 | cluster_0 | 1 | 302 |
| 11 | C8726 | cluster_0 | 2 | 301 |
| 12 | C9383 | cluster_0 | 2 | 301 |
| 13 | C8582 | cluster_3 | 18 | 285 |
| 14 | C10133 | cluster_3 | 16 | 287 |
| 15 | C8983 | cluster_0 | 1 | 302 |
| 16 | C3641 | cluster_1 | 3 | 300 |
| 17 | C5246 | cluster_0 | 1 | 302 |
| 18 | C10136 | cluster_0 | 1 | 302 |
| 19 | C7787 | cluster_0 | 1 | 302 |

Gambar 5. Hasil Clustering Data



Gambar 6. Proses Asosiasi FP-Growth

| Size | Support | Item 1 | Item 2 | Item 3 |
|------|---------|----------------------------|----------------------------|------------------|
| 1 | 0.994 | Whole Corn | | |
| 1 | 0.994 | Orange Flavored Fruit Bars | | |
| 1 | 0.970 | Hair Conditioner | | |
| 2 | 0.994 | Whole Corn | Orange Flavored Fruit Bars | |
| 2 | 0.970 | Whole Corn | Hair Conditioner | |
| 2 | 0.970 | Orange Flavored Fruit Bars | Hair Conditioner | |
| 3 | 0.970 | Whole Corn | Orange Flavored Fruit Bars | Hair Conditioner |

Gambar 7. Frequent Itemset beserta nilai Support

```

Association Rules
[Whole Corn] --> [Hair Conditioner] (confidence: 0.976)
[Orange Flavored Fruit Bars] --> [Hair Conditioner] (confidence: 0.976)
[Whole Corn] --> [Orange Flavored Fruit Bars, Hair Conditioner] (confidence: 0.976)
[Orange Flavored Fruit Bars] --> [Whole Corn, Hair Conditioner] (confidence: 0.976)
[Whole Corn, Orange Flavored Fruit Bars] --> [Hair Conditioner] (confidence: 0.976)
[Whole Corn] --> [Orange Flavored Fruit Bars] (confidence: 1.000)
[Orange Flavored Fruit Bars] --> [Whole Corn] (confidence: 1.000)
[Hair Conditioner] --> [Whole Corn] (confidence: 1.000)
[Hair Conditioner] --> [Orange Flavored Fruit Bars] (confidence: 1.000)
[Hair Conditioner] --> [Whole Corn, Orange Flavored Fruit Bars] (confidence: 1.000)
[Whole Corn, Hair Conditioner] --> [Orange Flavored Fruit Bars] (confidence: 1.000)
[Orange Flavored Fruit Bars, Hair Conditioner] --> [Whole Corn] (confidence: 1.000)
    
```

Gambar 8. Rule yang dihasilkan beserta nilai Confidence

| No. | Premises | Conclusion | Support | Confidence | Lift |
|-----|--|--|---------|------------|-------|
| 1 | Whole Com | Hair Conditioner | 0.970 | 0.976 | 1.006 |
| 2 | Orange Flavored Fruit Bars | Hair Conditioner | 0.970 | 0.976 | 1.006 |
| 3 | Whole Com | Orange Flavored Fruit Bars, Hair Conditioner | 0.970 | 0.976 | 1.006 |
| 4 | Orange Flavored Fruit Bars | Whole Com, Hair Conditioner | 0.970 | 0.976 | 1.006 |
| 5 | Whole Com, Orange Flavored Fruit Bars | Hair Conditioner | 0.970 | 0.976 | 1.006 |
| 6 | Whole Com | Orange Flavored Fruit Bars | 0.994 | 1 | 1.006 |
| 7 | Orange Flavored Fruit Bars | Whole Com | 0.970 | 1 | 1.006 |
| 8 | Hair Conditioner | Whole Com | 0.970 | 1 | 1.006 |
| 9 | Hair Conditioner | Orange Flavored Fruit Bars | 0.970 | 1 | 1.006 |
| 10 | Hair Conditioner | Whole Com, Orange Flavored Fruit Bars | 0.970 | 1 | 1.006 |
| 11 | Whole Com, Hair Conditioner | Orange Flavored Fruit Bars | 0.970 | 1 | 1.006 |
| 12 | Orange Flavored Fruit Bars, Hair Conditioner | Whole Com | 0.970 | 1 | 1.006 |

Gambar 9. Evaluasi dengan Support, Confidence, Lift

3.6. Pembangunan Sistem

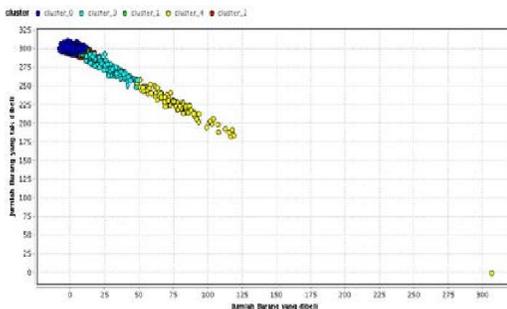
Pada tahapan ini dimulai dengan mendisain *input* dan *output* sistem dimana disain ini akan dipergunakan pada aplikasi rekomendasi produk. Sedangkan untuk pemrograman aplikasi rekomendasi produk ini menggunakan bahasa pemrograman *Java*.



Gambar 10. Tampilan Sistem Rekomendasi Promosi Produk

IV. HASIL PENELITIAN DAN PEMBAHASAN

4.1. Hasil Penelitian



Gambar 11. Grafik Hasil Clustering data Penjualan

Informasi yang bisa diperoleh dari hasil proses *clustering* tersebut adalah pada transaksi penjualan yang semula terdiri dari 1360 *record* transaksi, dapat dikelompokkan menjadi 5 *cluster*:

- 1) *Cluster* pertama terdiri dari 909 *record* transaksi, berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 1 - 2; dan jumlah barang yang tidak dibeli sekitar 301 - 302.
- 2) *Cluster* kedua terdiri dari 166 *record* transaksi, berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 3 - 7; dan jumlah barang yang tidak dibeli sekitar 296 - 300.
- 3) *Cluster* ketiga terdiri dari 66 *record* transaksi, berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 8 - 14; dan jumlah barang yang tidak dibeli sekitar 289 -295.
- 4) *Cluster* keempat terdiri dari 132 *record* transaksi, berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 15 -48; dan jumlah barang yang tidak dibeli sekitar 255 -288.
- 5) *Cluster* kelima terdiri dari 87 *record* transaksi, berisi transaksi yang memiliki jumlah barang yang dibeli sekitar 49 -303; dan jumlah barang yang tidak dibeli sekitar 0 -254.

Setelah proses *clustering* dilakukan, didapatkan 5 *cluster*. *Cluster* yang terbentuk dijadikan sebagai dataset baru yang lebih kecil daripada dataset sebelumnya. Kemudian setiap *cluster* akan dilakukan pendekatan asosiasi menggunakan algoritma *FP-Growth* untuk menganalisis keranjang pasar. Dengan dataset yang lebih kecil tujuan untuk meningkatkan akurasi dalam rekomendasi produk, sedangkan untuk perbandingan digunakan algoritma *apriori*.

Hasil akurasi untuk menentukan rekomendasi produk kepada pelanggan yang didapat dari penelitian ini dapat diukur dengan menggunakan *Lift Ratio*. *Lift Ratio* merupakan nilai yang menunjukkan kevalidan proses transaksi dan

memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B. Sebuah transaksi dikatakan valid jika mempunyai nilai *Lift Ratio* lebih dari 1, yang berarti bahwa dalam transaksi tersebut, produk A dan B benar-benar dibeli secara bersamaan. *Lift Ratio* mengukur seberapa penting *rule* yang telah terbentuk berdasarkan nilai *support* dan *confidence*.

Untuk eksperimen ditentukan nilai *Minimum Support* adalah 10% - 100% dan nilai *Minimum Confidence* 10% - 100%.

Berikut contoh *Rule* yang dihasilkan pada Cluster kelima dengan *Minimum Support* 50% dan *Minimum Confidence* 70% yang memiliki rata-rata *Lift Ratio* di atas 1 (lihat *Appendiks 2*).

4.2. Pembahasan

Penelitian ini membandingkan algoritma asosiasi antara *FP-Growth + K-Medoids*, *FP-Growth* dan *Apriori*. Untuk eksperimen ditentukan nilai *Minimum Support* adalah 10% - 100% dan nilai *Minimum Confidence* 10% - 100%. Dari hasil eksperimen yang dilakukan menunjukkan bahwa algoritma *FP-Growth + K-Medoids* lebih baik dalam menghasilkan *rule* dibandingkan *FP-Growth* ataupun *Apriori* yang tidak mampu menghasilkan *rule* pada dataset yang besar. Sedangkan untuk akurasi menggunakan *Lift Ratio*, algoritma *FP-Growth + K-Medoids* menghasilkan akurasi rata-rata lebih dari 1, yang mana sebuah transaksi dikatakan valid jika mempunyai nilai *Lift Ratio* lebih dari 1.

Berikut jumlah kevalidan *Rule* yang dihasilkan oleh Cluster 5. *Lift Ratio* yang lebih dari 1 dapat dikatakan memiliki tingkat kevalidan tinggi, sedangkan *Lift Ratio* yang kurang dari 1 dapat dikatakan memiliki tingkat kevalidan rendah (lihat *Appendiks 3*).

V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dari hasil penelitian ini menunjukkan bahwa penggunaan algoritma *K-Medoid* dapat mengkluster transaksi penjualan berdasarkan jumlah barang yang dibeli dan barang yang tidak dibeli menjadi 5 *cluster*, dan dari penggunaan algoritma *FP-Growth* dapat dihasilkan perbedaan asosiasi dan korelasi data pada masing-masing *cluster*, sedangkan untuk perbandingan digunakan algoritma *Apriori*. Dari hasil pengukuran menggunakan *Support*, *Confidence*, dan *Lift Ratio* didapati jumlah *rule* valid yang tinggi pada *cluster* kelima. Sehingga dapat diambil kesimpulan bahwa penerapan algoritma *K-Medoids* dan *FP-Growth* terbukti akurat untuk penentuan rekomendasi promosi produk pada dataset yang besar.

5.2. Saran

Penelitian ini telah menghasilkan suatu kaidah asosiasi yang akurat dengan jumlah kevalidan *rule* yang tinggi untuk rekomendasi promosi produk, namun untuk penelitian selanjutnya masih memerlukan pengembangan dalam beberapa hal, yakni:

- Menggunakan dataset yang lebih besar.
- Menggunakan dataset penjualan yang langsung didapat dari transaksi penjualan suatu toko.
- Menggunakan algoritma *clustering* dan algoritma asosiasi lain ataupun penggabungan beberapa algoritma lain.

UCAPAN TERIMA KASIH

Penelitian dengan judul “*K-MEDOIDS DAN FP-GROWTH* UNTUK ANALISIS KERANJANG PASAR” ini dapat penulis selesaikan sesuai rencana karena dukungan dari berbagai pihak yang

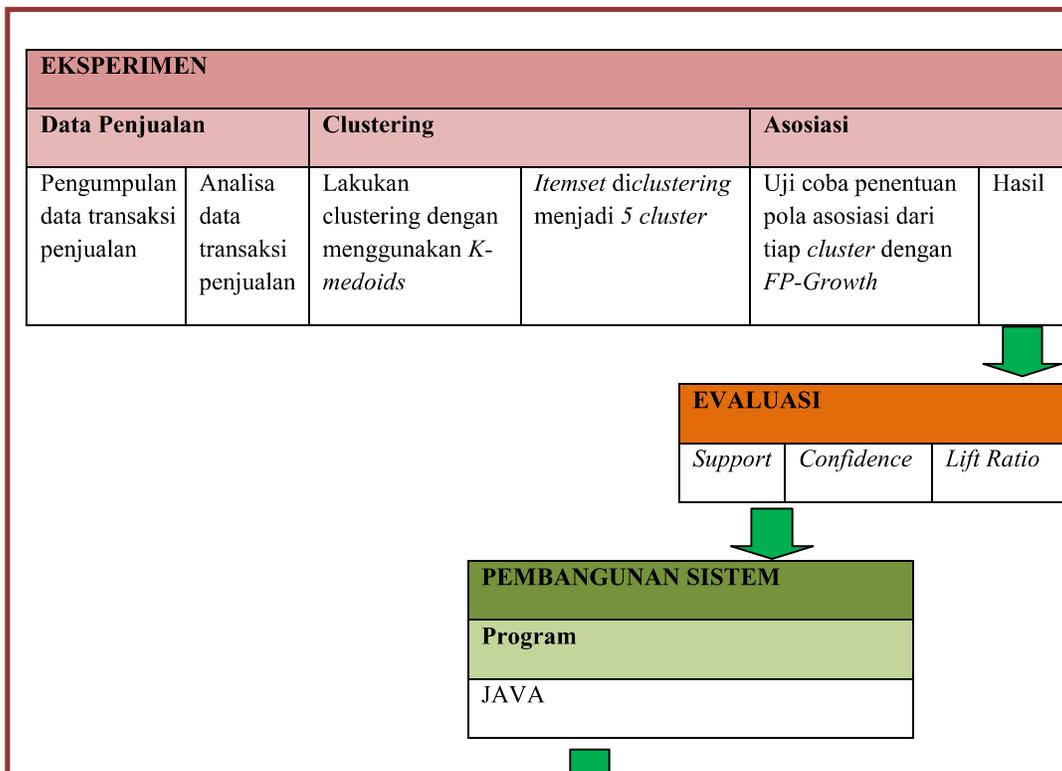
tidak ternilai besarnya terutama program pascasarjana MTI Universitas Dian Nuswantoro.

REFERENSI

- [1] H.K. Kim, J.K. Kim, and Q.Y. Chen, "A product network analysis for extending the market basket analysis," *Expert Systems With Applications*, vol. 39, 2012, pp. 7403-7410.
- [2] X. Wen-xiu, Q. Heng-nian, and H. Mei-li, "Market basket analysis based on text segmentation and association rule mining," *2010 First International Conference on Networking and Distributed Computing*, vol. 309313, 2010, pp. 309-313.
- [3] D.H. Setiabudi, G.S. Budhi, I.W. Purnama, and A. Noertjahyana, "Data Mining Market Basket Analysis ' Using Hybrid-Dimension Association Rules , Case Study in Minimarket X," *2011 International Conference on Uncertainty Reasoning and Knowledge Engineering*, 2011, pp. 196-199.
- [4] M. Ykhlef, "A Quantum Swarm Evolutionary Algorithm for mining association rules in large databases," *Journal of King Saud University - Computer and Information Sciences*, vol. 23, 2011, pp. 1-6.
- [5] M. Plasse, N. Niang, G. Saporta, A. Villeminot, and L. Leblond, "Combined use of association rules mining and clustering methods to find relevant links between binary rare attributes in a large data set," *Computational Statistics & Data Analysis*.
- [6] C. Wang, W. Lee, and C. Pang, "Applying Fuzzy FP-Growth to Mine Fuzzy Association Rules," *Engineering and Technology*, 2010.
- [7] K. Christidis, D. Apostolou, and G. Mentzas, "Exploring Customer Preferences with Probabilistic Topics Models," *Work*, 2010, pp. 1-13.
- [8] T. Velmurugan, "Efficiency of k-Means and K-Medoids Algorithms for Clustering Arbitrary Data Points," *International Journal Computer Technology & Applications*, vol. 3, 2012, pp. 1758-1764.
- [9] Y. Liu and Y. Guan, "FP-Growth Algorithm for Application in Research of Market Basket Analysis," *Design*, 2008.
- [10] L. XianWen and W. WeiQing, "Improved Algorithms Research for Association Rule Based on Matrix," *2010 International Conference on Intelligent Computing and Cognitive Informatics*, 2010, pp. 415-419.
- [11] Y. Liu and Y. Guan, "FP-Growth Algorithm for Application in Research of Market Basket Analysis," *IEEE 6th International Conference on Computational Cybernetics*, 2008, pp. 269-272.
- [12] B. Hoanca, "Using Market Basket Analysis to Estimate Potential Revenue Increases for a Small University Bookstore," *Information Systems*, vol. 1822, 2011, pp. 1-11.
- [12] D.T. Larose, *Discovering Knowledge In Data*, New Jersey: WILEY, 2005.
- [14] C. Vercellis, *Business Intelligence: Data Mining and Optimization for Decision Making*, Milan: WILEY, 2009.
- [15] G.S. Linoff and M.J. Berry, *Data Mining Techniques For Marketing, Sales, Customer Relationship Management*, WILEY, 2011.
- [16] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. Mclachlan, A. Ng, B. Liu, P.S. Yu, Z.Z. Michael, S. David, and J.H. Dan, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, 2008, pp. 1-37.
- [17] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2006.

APPENDIKS 1. Kerangka Pemikiran

| RUMUSAN MASALAH | | |
|--|--|---|
| Analisis Keranjang Pasar | Clustering | Asosiasi |
| Penentuan pola pembelian barang yang kurang akurat dapat mengakibatkan kebijakan rekomendasi produk dan promosi produk menjadi tidak tepat sasaran | Itemset yang besar tidak akan memberikan hasil yang maksimal pada asosiasi tetapi justru berakibat penentuan pola asosiasi yang kurang tepat | Aturan asosiasi cenderung mengabaikan <i>itemset</i> besar karena semua itemset berkontribusi terhadap asosiasi |



| TUJUAN | |
|---|--|
| Pengguna | IPTEK |
| Sebagai alat bantu pengambilan keputusan dalam membantu menentukan kebijakan, khususnya dalam bidang strategi pemasaran | Memberikan sumbangan bagi pengembangan teori yang berkaitan dengan proses pelatihan dari <i>FP-Growth</i> dan penerapan penggunaan <i>K-Medoids</i> sebagai metode <i>clustering</i> untuk menentukan rekomendasi produk.. |

APPENDIKS 2. Hasil Asosiasi dengan *Min-Support* 50% dan *Min-Confidence* 70%

| No | Premises | Conclusion | Support | Confidence | Lift |
|----|--------------------------------|---------------------|---------|------------|-------|
| 1 | <i>White Bread</i> | <i>Toothpaste</i> | 0,563 | 0,700 | 1,068 |
| 2 | <i>2pct, Milk</i> | <i>Potato Chips</i> | 0,563 | 0,700 | 0,998 |
| 3 | <i>White Bread</i> | <i>Potato Chips</i> | 0,575 | 0,714 | 1,019 |
| 4 | <i>Eggs</i> | <i>Potato Chips</i> | 0,563 | 0,742 | 1,059 |
| 5 | <i>White Bread</i> | <i>Eggs</i> | 0,621 | 0,771 | 1,017 |
| 6 | <i>2pct, Milk</i> | <i>Eggs</i> | 0,621 | 0,771 | 1,017 |
| 7 | <i>White Bread</i> | <i>2pct, Milk</i> | 0,632 | 0,786 | 0,977 |
| 8 | <i>2pct, Milk</i> | <i>White Bread</i> | 0,632 | 0,786 | 0,977 |
| 9 | <i>Toothpaste</i> | <i>Eggs</i> | 0,517 | 0,789 | 1,041 |
| 10 | <i>White Bread, 2pct, Milk</i> | <i>Eggs</i> | 0,506 | 0,800 | 1,055 |
| 11 | <i>Potato Chips</i> | <i>2pct, Milk</i> | 0,563 | 0,803 | 0,998 |
| 12 | <i>Potato Chips</i> | <i>Eggs</i> | 0,563 | 0,803 | 1,059 |
| 13 | <i>White Bread, Eggs</i> | <i>2pct, Milk</i> | 0,506 | 0,815 | 1,013 |
| 14 | <i>2pct, Milk, Eggs</i> | <i>White Bread</i> | 0,506 | 0,815 | 1,013 |
| 15 | <i>Eggs</i> | <i>White Bread</i> | 0,621 | 0,818 | 1,017 |
| 16 | <i>Eggs</i> | <i>2pct, Milk</i> | 0,621 | 0,818 | 1,017 |
| 17 | <i>Potato Chips</i> | <i>White Bread</i> | 0,575 | 0,820 | 1,019 |
| 18 | <i>Toothpaste</i> | <i>2pct, Milk</i> | 0,540 | 0,825 | 1,025 |
| 19 | <i>Wheat Bread</i> | <i>White Bread</i> | 0,506 | 0,846 | 1,052 |
| 20 | <i>Onions</i> | <i>2pct, Milk</i> | 0,529 | 0,852 | 1,059 |
| 21 | <i>Toothpaste</i> | <i>White Bread</i> | 0,563 | 0,860 | 1,068 |
| 22 | <i>Potatoes</i> | <i>White Bread</i> | 0,506 | 0,863 | 1,072 |

APPENDIKS 3. Jumlah Kevalidan *Lift Ratio*

| Min – Confidence | Min – Support (10%) | | Min – Support (20%) | | Min – Support (30%) | | Min – Support (40%) | | Min – Support (50%) | | Min – Support (60%) | | Min – Support (70%) | | Min – Support (80%) | | Min – Support (90%) | | Min – Support (100%) | | |
|------------------|---------------------|---------|---------------------|------|---------------------|----|---------------------|---|---------------------|---|---------------------|---|---------------------|---|---------------------|---|---------------------|---|----------------------|---|---|
| | T | R | T | R | T | R | T | R | T | R | T | R | T | R | T | R | T | R | T | R | |
| 10% | ~ | ~ | 49456 | 2808 | 21 | 17 | 26 | 2 | 2 | 8 | 4 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20% | ~ | ~ | 49456 | 2808 | 21 | 17 | 26 | 2 | 2 | 8 | 4 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30% | ~ | ~ | 46300 | 2445 | 21 | 17 | 26 | 2 | 2 | 8 | 4 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 40% | ~ | ~ | 37950 | 1898 | 21 | 16 | 26 | 2 | 2 | 8 | 4 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50% | ~ | ~ | 28873 | 1332 | 17 | 12 | 26 | 2 | 2 | 8 | 4 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60% | ~ | ~ | 20030 | 746 | 12 | | 20 | 1 | 2 | | | | | | | | | | | | |
| 70% | 2463294 | 1220846 | 12371 | 539 | 76 | 13 | 1 | 1 | | | | | | | | | | | | | |
| 80% | 1643867 | 821930 | 6874 | 63 | 42 | | | | | | | | | | | | | | | | |
| 90% | 1476172 | 165861 | 2474 | 0 | 3 | 12 | 78 | 3 | 3 | 1 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 100% | 550926 | 78703 | 261 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Keterangan:

T → Tinggi, R → Rendah