

## TINJAUAN PUSTAKA

### 2.1 Analisis Deskriptif

Analisis deskriptif adalah metode penelitian dengan cara mengumpulkan data-data sesuai dengan yang sebenarnya kemudian data-data tersebut disusun, diolah dan dianalisis untuk dapat memberikan gambaran mengenai masalah yang ada. Pada analisis deskriptif data biasanya ditampilkan dalam bentuk tabel biasa atau tabel frekuensi, grafik, diagram batang, diagram garis, diagram lingkaran, ukuran pemusatan data, ukuran penyebaran data dan sebagainya (Sugiyono, 2010).

### 2.2 Bisnis Ritel

Ritel berasal dari kata *retail* yang berarti eceran. Bisnis ritel merupakan suatu bisnis menjual produk dan jasa pelayanan yang telah diberi nilai tambah untuk memenuhi kebutuhan pribadi, keluarga, atau pengguna akhir lainnya (Sunarto, 2010). Kegiatan ritel (*retailing*) merupakan salah satu industri yang terbesar di dunia karena melibatkan peran berbagai golongan peritel mulai dari skala pengecer hingga *Hypermarket*. Dalam sebuah negara, kegiatan ritel menjadi penting guna memacu produktivitas, daya beli masyarakat dan produsen sehingga dapat menggerakkan roda perekonomian (Setiawan, 2009).

Ritel meliputi semua perusahaan yang memiliki aktivitas penjualan barang atau jasa secara langsung kepada konsumen akhir, umumnya tidak melakukan transformasi produk, dan memberikan pelayanan akan penjualan

produk (Bell, 2004). Proses ritel sendiri merupakan langkah akhir akan distribusi produk, dimana *retailer* berperan untuk menjual produk dalam kuantitas kecil kepada masyarakat. Pelayanan yang diberikan umumnya meliputi transportasi dan *stock keeping* untuk meyakinkan bahwa produk tersedia ketika dibutuhkan.

Dalam menjalankan perannya, peritel menjalankan fungsi dalam meningkatkan nilai produk ataupun jasa yang dijual dengan berbagai cara seperti berikut:

1. Peyedia variasi

Dalam upaya memenuhi kepuasan pelanggan, peritel menyediakan variasi produk. Umumnya toko serba ada (*Supermarket*) yang berskala menengah menjual 20.000 ribu hingga 40.000 produk. Hal ini menyebabkan pelanggan memiliki berbagai pilihan dari segi merek, bentuk, ukuran, warna, serta harga dalam satu lokasi.

2. Pemecah volume

Dalam upaya mengurangi biaya distribusi produk, produsen atau penggrosir (*wholesaler*) biasanya mengirimkan produk dalam jumlah besar kepada peritel. Selanjutnya peritel akan menjual produk dalam jumlah yang lebih kecil yang disesuaikan dengan pola konsumsi konsumen. Hal ini menunjukkan fungsi peritel dalam memecah volume (*breaking bulk*).

### 3. Pengadaan (*Procurement*)

Fungsi terpenting peritel adalah saat menyimpan persediaan. Peritel akan memastikan produk tersedia ketika pelanggan membutuhkannya. Hal ini menunjukkan bahwa peritel mengambil alih biaya *inventory holding cost* untuk kemudian ditukar dengan *margin profit* sebagai *trade off* kepada konsumen atas minimalisasi biaya penyimpanan.

### 4. Penyedia layanan

Peritel tidak hanya menjual produk tetapi juga menyediakan layanan berkenaan dengan produk yang dijualnya. Peritel menyediakan pramuniaga untuk membantu pelanggan dalam berbelanja, melakukan pengemasan, bahkan pengantaran produk ke kendaraan konsumen.

Saat ini tren bisnis ritel telah bergeser dari awalnya fokus akan pemilihan dan pembelian produk guna mengakomodasi variasi produk menjadi manajemen yang mengutamakan pendekatan personal. Konsep ini merupakan pendekatan holistik akan manajemen dan marketing guna meningkatkan kesempatan untuk menjadi *consumer oriented*, meningkatkan kontak personal dengan konsumen, mengumpulkan informasi akan kebiasaan konsumen, dan menyesuaikan organisasi sesuai dengan perilaku dan preferensi pelanggan (Mulhern, 1996).

Sejak diperkenalkan pertama kali di Indonesia pada era 1970-an, saat ini terdapat tiga jenis pasar ritel yang berkembang pesat, yaitu *Minimarket*, *Supermarket*, dan *hypermarket*. Perbedaan utama dari ketiganya terletak pada

luas lahan usaha dan *range* jenis barang yang diperdagangkan. Karakteristik dari ketiga jenis pasar tersebut adalah sebagai berikut (Pandin, 2009).

**Tabel 2.1** Karakteristik Pasar Ritel Modern

Uraian	Minimarket	Supermarket	Hypermarket
Barang yang diperdagangkan	Berbagai macam kebutuhan rumah tangga termasuk kebutuhan sehari-hari	Berbagai macam kebutuhan rumah tangga termasuk kebutuhan sehari-hari	Berbagai macam kebutuhan rumah tangga termasuk kebutuhan sehari-hari
Jumlah item	< 5000 item	5000 – 25000 item	> 25000 item
Jenis produk	Makanan kemasan, barang-barang higienis produk	Makanan, barang-barang rumah tangga	Makanan, barang-barang rumah tangga, elektronik, pakaian, alat olahraga
Uraian	Minimarket	Supermarket	Hypermarket
Model penjualan	Dilakukan secara eceran, langsung pada konsumen akhir dengan cara swalayan	Dilakukan secara eceran, langsung pada konsumen akhir dengan cara swalayan	Dilakukan secara eceran, langsung pada konsumen akhir dengan cara swalayan
Luas lantai usaha (Berdasar perpres no.112 th 2007)	Maksimal 400 m <sup>2</sup>	4000 – 5000 m <sup>2</sup>	> 5000 m <sup>2</sup>
Luas lahan parkir	Minim	Standart	Sangat luas
Modal (di luar tanah bangunan)	s/d 200 juta	200 juta – 10 miliar	> 10 miliar

### 2.3 Perilaku Konsumen (*Consumer Behavior*)

Perilaku konsumen adalah tindakan-tindakan yang dilakukan oleh individu, kelompok atau organisasi yang berhubungan dengan proses pengambilan keputusan dalam mendapatkan, menggunakan barang-barang atau jasa ekonomis yang dapat dipengaruhi lingkungan (Mangkunegara & Prabu, 2009). David dan Bitta (Loudon & Bitta, 1988) dalam bukunya lebih menekankan perilaku konsumen sebagai suatu proses pengambilan

keputusan. Mereka mengatakan bahwa perilaku konsumen adalah proses pengambilan keputusan yang mensyaratkan aktivitas individu untuk mengevaluasi, memperoleh, menggunakan atau mengatur barang dan jasa.

Menurut Setiadi (Setiadi & Nugroho, 2003) banyak faktor yang mempengaruhi perilaku konsumen dalam membuat keputusan pembelian. Faktor tersebut sebagian besar tidak dapat dikendalikan oleh pemasaran, tetapi harus diperhatikan, namun secara garis besar factor-faktor yang mempengaruhi perilaku konsumen dibagi menjadi dua bagian yaitu eksternal dan faktor internal. Faktor internal yang mempengaruhi perilaku konsumen seperti: motivasi, kepribadian, pembelanjaan, sikap, persepsi konsumen. Sedangkan faktor eksternal meliputi: kelompok rujukan, kelas sosial, budaya dan komunikasi.

Dari berbagai definisi tersebut, dapat ditarik kesimpulan bahwa perilaku konsumen menyoroti perilaku baik individu maupun rumah tangga, perilaku konsumen menyangkut suatu proses pengambilan keputusan sebelum pembelian sampai dengan mengkonsumsi produk, dan tujuan mempelajari perilaku konsumen adalah untuk menyusun strategi pemasaran yang berhasil.

#### 2.4 *Preprocessing*

*Database* yang ada sekarang ini kebanyakan merupakan data kotor, kebanyakan masih terdapat *misssing* data, *noise*, dan data yang tidak konsisten didalamnya. Kualitas data yang rendah ini akan menghasilkan hasil yang berkualitas rendah juga. Oleh karena itu perlu adanya *preposessing* untuk meningkatkan efisiensi pada proses yang dilakukan (Han & Kember, 2006). Saat menggunakan data *warehouse* untuk melakukan analisis, maka data yang digunakan akan sangat besar. Menganalisis data yang kompleks dengan jumlah yang besar akan

membutuhkan waktu yang lama. Hal ini sangat tidak praktis dan tidak efisien. Teknik reduksi data dapat diaplikasikan untuk mengurangi volume data. Reduksi data bisa digunakan untuk mengurangi jumlah variabel ataupun jumlah objek penelitian. Pada kasus ini reduksi data yang digunakan bertujuan untuk mengurangi jumlah variabel penelitian. Proses reduksi dimensi yang dilakukan pada data transaksi dilakukan pada saat pemindaian pertama. Pada pemindaian pertama, semua item yang muncul pada transaksi yang lebih kecil dari 100 teratas yang ditetapkan disebut sebagai *infrequent item*. Selanjutnya *infrequent* data akan dibuang dari transaksi sehingga transaksi tersebut tidak akan menjadi bagian dari *itemset*. Item dalam setiap transaksi akan diurutkan secara menurun berdasarkan frekuensi dalam *database*.

## 2.5 Aturan Asosiasi (*Association Rules*)

Menurut Gemala (Gemala, Arifin, & Aliffianto, 2011) *Association rules* merupakan sebuah aturan tertentu atau *rules* yang menyatakan sebuah hubungan korelasi antara tingkat kemunculan beberapa atribut dalam sebuah *database*. *Association Rule* adalah bentuk jika “kejadian sebelumnya” kemudian “konsekuensinya” (*IF antecedent, THEN consequent*). Bentuk umum dari *Association Rules* adalah:

$$A_1 \dots \dots, A_n \rightarrow B \quad 2.1)$$

Yang artinya bahwa pelanggan yang membeli produk A juga mempunyai peluang yang cukup besar untuk membeli produk B, dimana tidak ada batasan pada jumlah dari *item-item* pada *head* ataupun *body* dari sebuah *rule*. Bentuk lain yang lebih kompleks dari *Association Rule* adalah:

$$A, C \rightarrow B, D \quad 2.2)$$

Yang artinya bahwa pelanggan yang membeli produk A dan C juga mempunyai kecenderungan untuk membeli produk B dan D. Salah satu masalah dalam pencarian *Association Rule* adalah sangat banyaknya kemungkinan ditemukannya *rules* yang belum tentu merupakan *rules* yang baik dan dapat dipercaya. Untuk mengetahui mana *rule* yang baik maka

terdapat suatu parameter untuk menilai kualitas dari *association rule* yang dihasilkan, yaitu *support* dan *confidence* dan *improvement / lift ratio* (Firdhana, 2006).

Analisis asosiasi didefinisikan sebagai suatu proses untuk menemukan semua aturan asosiasi yang memenuhi syarat minimum untuk *support* (minimum *support*) dan syarat minimum untuk *confidence* (minimum *confidence*). Jika *supportnya*  $\geq$  minimum *support* dan *confidencenya*  $\geq$  minimum *confidence*, maka *rule* tersebut bisa dikatakan sebagai *interesting rule*. Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam *database*. Nilai *support* sebuah *item* diperoleh dengan perhitungan seperti berikut.

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total jumbuh transaksi keseluruhan}} \times 100\% \quad 2.3)$$

Sementara itu, nilai *support* dari 2 item diperoleh dari perhitungan berikut:

$$\text{Support (A} \cap \text{B)} = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Total jumbuh transaksi keseluruhan}} \times 100\% \quad 2.4)$$

Setelah semua pola frekuensi ditemukan, barulah dicari aturan asosiasinya yang memenuhi syarat minimum untuk *confidence* dengan perhitungan *confidence* sebagai berikut:

$$\text{Confidence (A} \cap \text{B)} = \frac{\text{Support (A} \cap \text{B)}}{P(A)} \times 100\% \quad 2.5)$$

*Lift ratio* adalah parameter penting selain *support* dan *confidence* dalam *association rule*. *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. Sebuah transaksi dikatakan valid jika mempunyai nilai *Lift / improvement* lebih  $>1$  atau  $=1$ , yang berarti bahwa dalam transaksi tersebut, produk A dan B benar-benar dibeli secara bersamaan (Kuswayati & Tjahyadi, 2015). Nilai *lift ratio* sebuah *item* diperoleh dengan perhitungan seperti berikut:

$$Lift\ Ratio = \frac{Support\ (A \cap B)}{Support\ (A) \times Support\ B} \times 100\% \quad 2.6)$$

Menurut ngatimin salah satu contoh penerapan *Association Rule* adalah *Market Basket Analysis*. *Association Rule* menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan, sehingga *Association Rule* juga sering disebut dengan istilah *Market Basket Analysis*.

## 2.6 Ukuran Kepercayaan Rule (*Interesting Measure*)

Menurut (Han, Kamber, & Pei, 2012), terdapat dua ukuran kepercayaan yang menunjukkan kepastian dan tingkat kegunaan suatu *rule* yang ditemukan yaitu:

### 1. *Support*

*Support* (dukungan) merupakan suatu ukuran yang menunjukkan seberapa besar dominasi suatu *item* dan *itemset* dari keseluruhan transaksi.

### 2. *Confidence*

*Confidence* (tingkat kepercayaan) adalah suatu ukuran yang menunjukkan hubungan antar *item* secara *conditional* (misalnya seberapa sering *item* dibeli jika orang membeli *item* A).



Contoh:

Beli (x, “Kopi”)  $\rightarrow$  Beli (x, “Gula”) [*Support* = 50%; *Confidence* = 80%]

Keterangan:

Kopi dan Gula dibeli bersamaan sebesar 50% dari seluruh transaksi dan 80% dari semua konsumen yang membeli kopi juga membeli gula.

## 2.7 Analisis Keranjang Pasar (*Market Basket Analysis*)

Analisis keranjang pasar (*Market Basket Analysis*) adalah salah satu teknik pemodelan dalam data mining berdasarkan teori yang mana jika anda membeli suatu grup item, anda akan memiliki kemungkinan membeli itemset yang lain (Han et al., 2012). Dalam hal ini, kuantitas dari sebuah item yang dibeli konsumen tidak mempengaruhi proses analisis ini. Market Basket Analysis hanya berdasarkan tipe-tipe item yang berbeda, tidak peduli seberapa banyak kuantitasnya (Fatihatul, Setiawan, & Rosadi, 2011).

Proses ini menganalisis *buying habits* dari para konsumen dengan menemukan hubungan asosiasi antar *item-item* yang berbeda yang seringkali dibeli oleh konsumen. Hasil dari proses analisis ini nantinya akan sangat berguna bagi perusahaan ritel khususnya seperti *Minimarket* dan *Supermarket* untuk mengembangkan strategi pemasaran dan proses pengambilan keputusan dengan melihat *item-item* berbeda yang sering dibeli secara bersamaan oleh konsumen. Berikut merupakan beberapa kegunaan lain hasil analisa keranjang pasar oleh peritel.

1. Mengembangkan strategi periklanan dan promosi yang lebih *profitable*

Peritel menggunakan analisa keranjang pasar untuk membuat iklan dan promosi lebih terukur dan terarah. Hasil analisa keranjang pasar dapat membantu peritel untuk melihat secara lebih teliti apakah iklan dan promosi yang dilakukan hanyamerubah pendapatan (*shifting revenue*) atau benar- benar menaikkan penjualan atau adakah produk lain yang dikorbankan akibat iklan dan promosi salah satu produk.

2. *Targeting* yang lebih terarah

Analisa keranjang pasar digunakan untuk mengoptimasi kampanye dan promosi untuk meningkatkan marjin dan tingkat penjualan pada target yang lebih tepat.

3. Promosi yang lebih baik dengan promosi kartu berlangganan

Peritel dapat menggunakan kartu berlangganan konsumen untuk melihat karakteristik perilaku pembelian konsumen beserta *lifecycle* pembelajarannya.

4. Meningkatkan daya tarik toko

Hasil analisa keranjang pasar dapat digunakan untuk mengidentifikasi produk- produk strategis yang dapat membuat konsumen setia dan tetap kembali ke toko.

5. Meningkatkan ukuran dan nilai dari keranjang belanja

Dengan menggunakan kartu berlangganan, peritel dapat mengetahui kuantitas kedatangan konsumen dan isi dari keranjang belanja konsumen. Hal ini dapat dijadikan dasar dalam menetapkan strategi lebih lanjut.

6. Mempelajari dan menggunakan pasar sebagai bahan uji coba

Peritel menggunakan analisa keranjang pasar untuk menentukan bagaimana cara meningkatkan rata-rata nilai keranjang belanja tanpa mengorbankan margin produk.

7. Menentukan *magic price*

Apabila dibandingkan dengan metode tradisional, analisa keranjang pasar lebih efektif dan efisien untuk menentukan *pricing zone* (batas kenaikan harga sampai konsumen berhenti membeli produk karena perbedaan harga dibandingkan kompetitor terdekat ketika peritel ingin menaikkan harga.

8. Menyesuaikan tingkat *inventory* sesuai kebutuhan

Penyesuaian tingkat *inventory* dapat dilakukan dengan pengaturan toko dan pemilihan produk yang dijual sesuai lokasi demografis. Hal ini dilakukan juga dengan pertimbangan analisa keranjang pasar akan tingkat penjualan produk.

9. Mengidentifikasi produk yang memiliki kinerja penjualan serupa

Pengolahan data dengan analisa keranjang pasar juga dapat menunjukkan rekapitulasi kinerja penjualan antar produk sehingga dapat diketahui produk yang tergolong *fast moving* dan produk *slow moving* untuk dianalisa mengenai strategi promosi dan penjualannya.

10. Menentukan produk yang harus di *cross-selling*

Hasil dari analisa keranjang pasar akan menampilkan produk terlaris dan produk yang kurang baik penjualannya. Berdasarkan hal ini dapat dilakukan strategi *cross selling* untuk mengangkat penjualan produk yang kurang diminati.

11. Mengidentifikasi produk yang sukses tanpa pengaruh yang signifikan dari elemen lain

Analisa keranjang pasar selain menghasilkan rekapitulasi produk, juga dihasilkan keterangan mengenai produk lain yang kerap kali menyertai pembelian suatu produk tertentu. Produk yang banyak dibeli dengan minimum produk yang menyertainya dapat dikatakan produk yang *independent* sehingga penempatannya dapat lebih fleksibel

12. Menganalisa produk yang harus diletakkan berdekatan

Hasil awal dari analisa keranjang pasar adalah pola asosiasi produk yang dapat langsung diinterpretasikan untuk menganalisa produk yang harus diletakkan berdekatan karena sering dibeli bersamaan.

13. Mengoptimasi *layout* toko

Hasil analisa keranjang pasar yang berupa hubungan asosiasi produk dapat diimplementasikan pada *layout* toko untuk mengoptimasi *layout* untuk meningkatkan penjualan produk.

## 2.8 Algoritma *FP-Growth*

### 2.6.1 Pengertian *Frequent Pattern Growth (FP-Growth)*

Algoritma *FP-Growth* merupakan salah satu alternatif algoritma yang cukup efektif untuk mencari himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data yang besar. Algoritma *FP-Growth* merupakan algoritma *Association Rules* yang cukup sering dipakai dan merupakan pengembangan dari algoritma Apriori. Algoritma Apriori menghasilkan kombinasi yang sangat banyak sehingga data yang dihasilkan tidak efisien (Gunadi & Sensuse, 2012).

Algoritma ini hanya melakukan 2 kali proses *scanning database* untuk menentukan *frequent itemset* dan juga tidak membutuhkan *generate candidate* seperti yang dilakukan pada algoritma Apriori. Struktur data yang digunakan untuk mencari *frequent itemset* dengan algoritma *FP-Growth* adalah perluasan dari penggunaan sebuah pohon *prefix*, yang biasa disebut adalah *FP-Tree*. Dengan menggunakan *FP-Tree*, algoritma *FP-Growth* dapat langsung mengekstrak *frequent Itemset* dari *FP-Tree* yang telah terbentuk sehingga pemrosesanpun data bisa lebih cepat.

### 2.6.2 Pembangunan *FP-Tree*

*FP-Tree* merupakan struktur penyimpanan data yang dimampatkan. *FP-Tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam *FP-Tree*. Karena dalam setiap transaksi yang dipetakan, mungkin transaksi yang memiliki *item* yang sama, maka lintasannya memungkinkan untuk saling menimpa. Semakin banyak data transaksi yang memiliki *item* yang sama, maka proses pemampatan dengan

struktur data *FP-Tree* semakin efektif. Kelebihan dari *FP-Tree* adalah hanya memerlukan dua kali pemindaian data transaksi yang terbukti sangat efisien.

Misal  $I = \{a_1, a_2, \dots, a_n\}$  adalah kumpulan dari *item*. Dan basis data transaksi  $DB = \{T_1, T_2, \dots, T_n\}$ , di mana  $T_i$  ( $i \in [1..n]$ ) adalah sekumpulan transaksi yang mengandung *item* di  $I$ . Sedangkan *support* adalah penghitung (*counter*) frekuensi kemunculan transaksi yang mengandung suatu pola. Suatu pola dikatakan sering muncul (*frequent pattern*) apabila *support* dari pola tersebut tidak kurang dari suatu konstanta  $\xi$  (batas ambang *minimum support*) yang telah didefinisikan sebelumnya. Permasalahan mencari pola *frequent* dengan batas ambang *minimum support count*  $\xi$  inilah yang dicoba untuk dipecahkan oleh *FP-Growth* dengan bantuan Struktur *FP-Tree*. Adapun *FP-Tree* adalah sebuah pohon dengan definisi sebagai berikut:

- *FP-Tree* dibentuk oleh sebuah akar yang diberi label *null*, sekumpulan upapohon yang beranggotakan *item-item* tertentu, dan sebuah tabel *frequent header*.
- Setiap simpul dalam *FP-Tree* mengandung tiga informasi penting, yaitu label *item*, menginformasikan jenis *item* yang direpresentasikan simpul tersebut, *support count*, merepresentasikan jumlah lintasan transaksi yang melalui simpul tersebut, dan *pointer* penghubung yang menghubungkan simpul-simpul dengan label *item* sama antar-lintasan, ditandai dengan garis panah putus-putus.

Contoh 1 Misalkan diberikan tabel data transaksi sebagai berikut, dengan *minimum support count*  $\xi = 2$

**Tabel 2.2** Data Transaksi Mentah

No	Transaksi
1	a, b
2	b, c, d, g, h
3	a, c, d, e, f
4	a, d, e
5	a, b, z, c
6	a, b, c, d
7	a, r
8	a, b, c
9	a, b, d
10	b, c, e

Frekuensi kemunculan tiap *item* dapat dilihat pada tabel berikut:

**Tabel 2.3** Frekuensi Kemunculan Tiap Karakter

Item	Frekuensi
A	8
B	7
C	6
D	5
E	3

F	1
R	1
Z	1
G	1
H	1

Setelah dilakukan pemindaian pertama didapat *item* yang memiliki frekuensi di atas *support count*  $\xi=2$  adalah a, b, c, d, dan e. Kelima *item* inilah yang akan berpengaruh dan akan dimasukkan ke dalam *FP-Tree*, selebihnya (r, z, g, dan h) dapat dibuang karena tidak berpengaruh signifikan.

Tabel berikut mendata kemunculan *item* yang *frequent* dalam setiap transaksi, diurut berdasarkan yang frekuensinya paling tinggi.

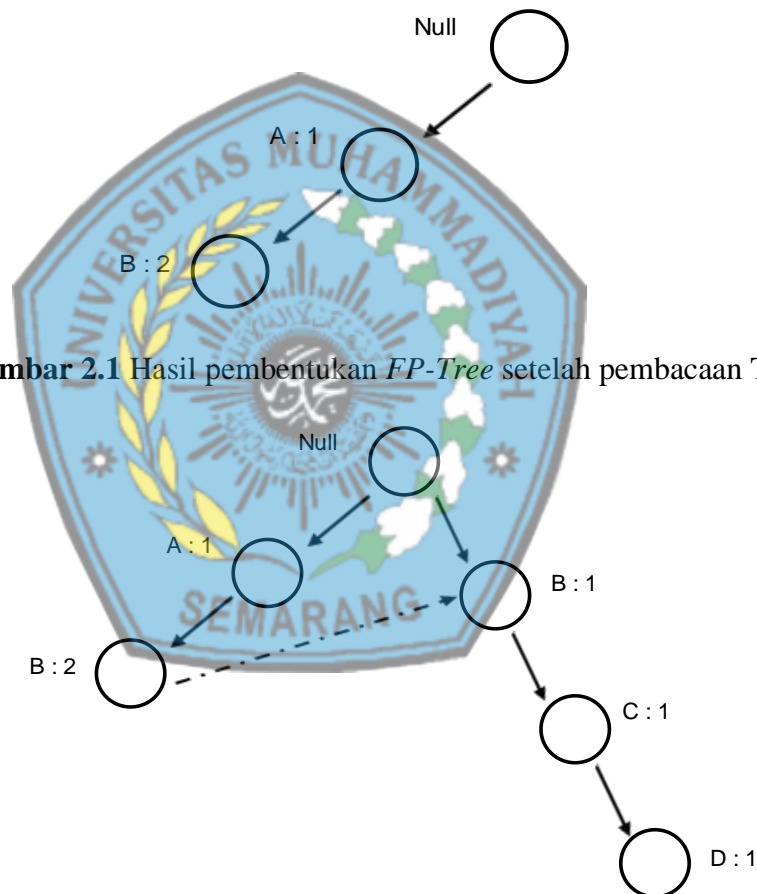
**Tabel 2.4** Data Transaksi

TID	Item
1	a, b
2	b, c, d
3	a, c, d, e
4	a, d, e
5	a, b, c
6	a, b, c, d



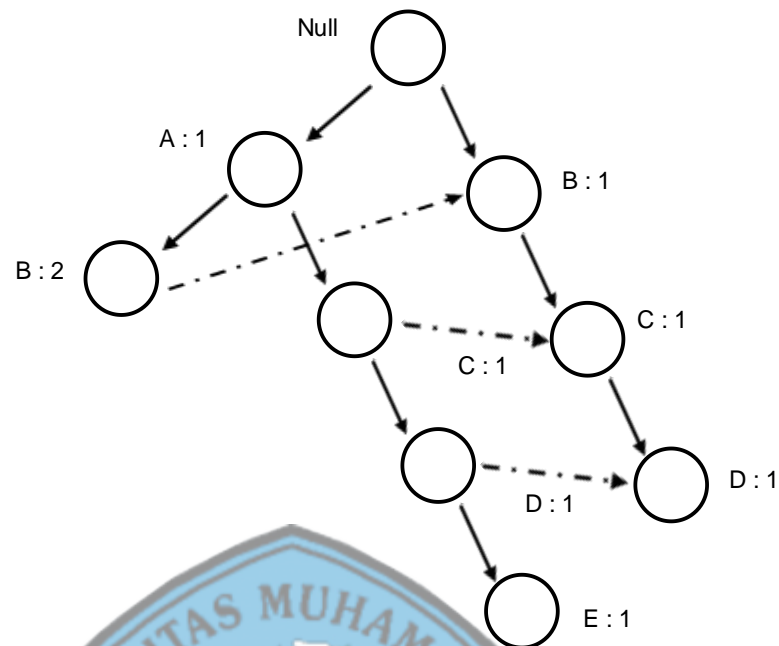
7	a
8	a, b, c
9	a, b, d
10	b, c, e

Gambar di bawah ini memberikan ilustrasi mengenai pembentukan *FP-Tree* setelah pembacaan TID 1.

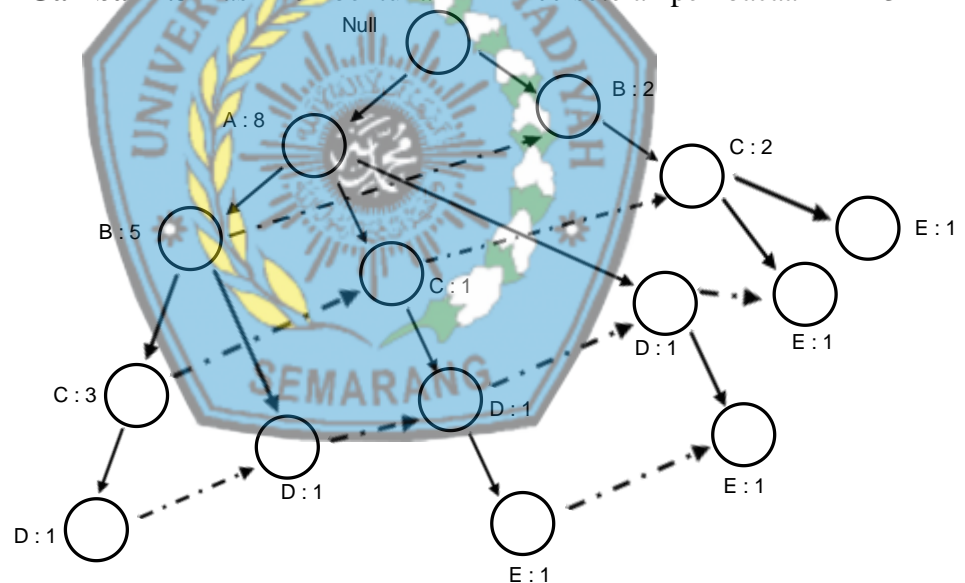


**Gambar 2.1** Hasil pembentukan *FP-Tree* setelah pembacaan TID 1

**Gambar 2.2** Hasil Pembentukan *FP-Tree* setelah pembacaan TID 2



**Gambar 2.3** Hasil Pembentukan *FP-Tree* setelah pembacaan TID 3



**Gambar 2.4** Hasil Pembentukan *FP-Tree* setelah pembacaan TID 10  
Diberikan 10 data transaksi dengan 5 jenis *item* seperti pada tabel di

atas. Gambar 1 – 4 menunjukkan proses terbentuknya *FP-Tree* setiap TID dibaca. Setiap simpul pada *FP-Tree* mengandung nama sebuah *item* dan *counter support* yang berfungsi untuk menghitung frekuensi kemunculan *item* tersebut dalam tiap lintasan transaksi.

*FP-Tree* yang merepresentasikan data transaksi pada Tabel 2.4 dibentuk dengan cara sebagai berikut:

1. Kumpulan data dipindai pertama kali untuk menentukan *support count* dari setiap *item*. *Item* yang tidak *frequent* dibuang, sedangkan *frequent item* dimasukkan dan disusun dengan urutan menurun, seperti yang terlihat pada Tabel 2.4.
2. Pemindaian kedua, yaitu pembacaan TID pertama {a,b} akan membuat simpul a dan b, sehingga terbentuk lintasan transaksi  $\text{Null} \rightarrow a \rightarrow b$ . Support count dari setiap simpul bernilai awal
3. Setelah pembacaan transaksi kedua {b,c,d}, terbentuk lintasan kedua yaitu  $\text{Null} \rightarrow b \rightarrow c \rightarrow d$ . Support count masing-masing count juga bernilai awal 1. Walaupun b ada pada transaksi pertama, namun karena prefix transaksinya tidak sama, maka transaksi kedua ini tidak bisa dimampatkan dalam satu lintasan.
4. Transaksi keempat memiliki prefix transaksi yang sama dengan transaksi pertama, yaitu a, maka lintasan transaksi ketiga dapat ditimpakan di a, sambil menambah support count dari a, dan selanjutnya membuat lintasan baru sesuai dengan transaksi ketiga. (lihat Gambar 2.3)
5. Proses ini dilanjutkan sampai *FP-Tree* berhasil dibangun berdasarkan tabel data transaksi yang diberikan.

### 2.6.3 Penerapan Algoritma FP-Growth

Setelah tahap pembangunan *FP-Tree* dari sekumpulan data transaksi, akan diterapkan algoritma *FP-Growth* untuk mencari *frequent itemset* yang signifikan. Algoritma *FP-Growth* dibagi menjadi tiga langkah utama, yaitu (Han et al., 2012):

1. Tahap Pembangkitan *Conditional Pattern Base*

*Conditional Pattern Base* merupakan *subdatabase* yang berisi *prefix path* (lintasan prefix) dan *suffix pattern* (pola akhiran). Pembangkitan *conditional pattern base* didapatkan melalui *FP-Tree* yang telah dibangun sebelumnya.

2. Tahap Pembangkitan *Conditional FP-tree*

Pada tahap ini, *support count* dari setiap *item* pada setiap *conditional pattern base* dijumlahkan, lalu setiap *item* yang memiliki jumlah *support count* lebih besar sama dengan minimum *support count* akan dibangkitkan dengan *conditional FP-Tree*.

3. Tahap Pencarian *frequent itemset*

Apabila *conditional FP-Tree* merupakan lintasan tunggal (*single path*), maka didapatkan *frequent itemset* dengan melakukan kombinasi *item* untuk setiap *conditional FP-Tree*. Jika bukan lintasan tunggal, maka dilakukan pembangkitan *FP-Growth* secara rekursif.