BAB II

LANDASAN TEORI

2.1 *Data Mining*

2.1.1 Pengertian *Data Mining*

Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, *data mining* dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field* dari sebuah relasional *database* yang besar. [1]

Kemampuan *Data mining* untuk mencari informasi bisnis yang berharga dari basis data yang sangat besar, dapat dianalogikan dengan penambangan logam mulia dari lahan sumbernya, teknologi ini dipakai untuk :

1. Prediksi trend dan sifat-sifat bisnis, dimana *data mining* mengotomatisasi proses pencarian informasi pemprediksi di dalam basis data yang besar.
2. Penemuan pola-pola yang tidak diketahui sebelumnya, dimana data mining "menyapu" basis data, kemudian mengidentifikasi pola-pola yang sebelumnya tersembunyi dalam satu sapuan.
3. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

2.1.2 Fungsi *Data Mining*

Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna serta meningkatkan pengetahuan bagi pengguna. Pada dasarnya, *data mining* mempunyai empat fungsi dasar yaitu :

1. Fungsi Klasifikasi (*Classification)*

   Data mining dapat digunakan untuk mengelompokkan data-data yang jumlahnya besar menjadi data-data yang lebih kecil.

2. Fungsi Segmentasi (*Segmentation)*

   Data mining dapat digunakan untuk melakukan segmentasi (pembagian) tehadap data berdasarkan karakteristik tertentu.

3. Fungsi Asosiasi (*Association*)

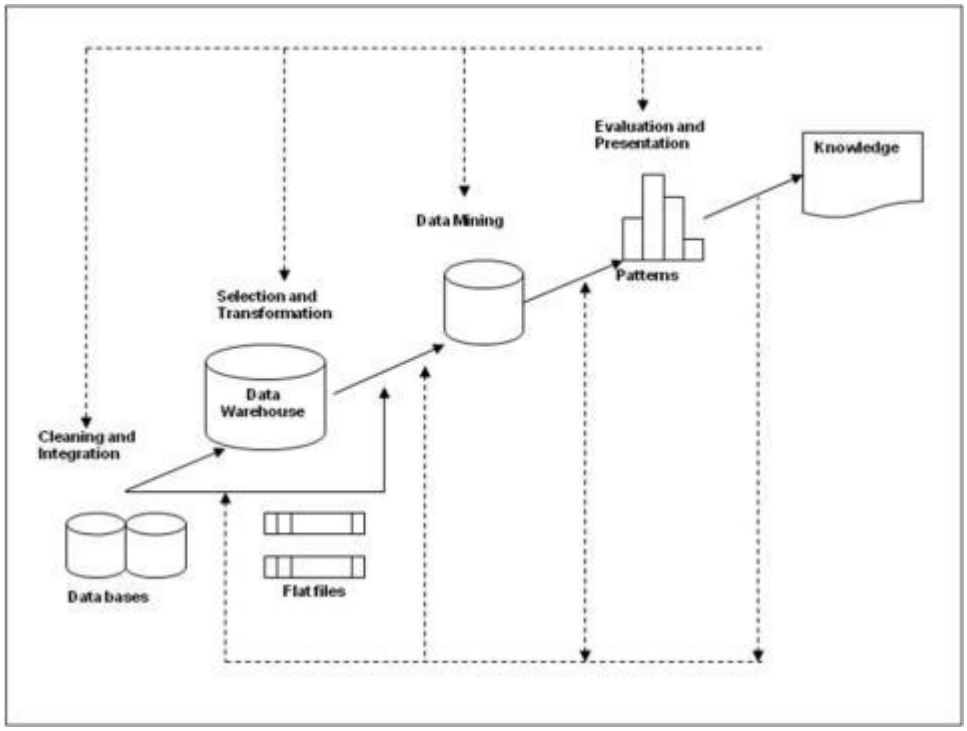
   Data mining digunakan untuk mencari hubungan antara karakteristik tertentu.

4. Fungsi Pengurutan (*Sequencing*)

   Data mining digunakan untuk mengidentifikasikan perubahan pola yang telah terjadi dalam jangka waktu tertentu.

2.2 Metode *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Database adalah proses untuk menggali dan menganalisis sejumlah data dan mengestrak informasi dan pengetahuian yang berguna.Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan untuk keperluan pengambilan keputusan. Proses dalam KDD adalah proses yang digambarkan pada dan terdiri dari rangkaian proses sebagai berikut :



Gambar 2.1 Tahapan *Knowledge Discovery in Database*
Sumber : (Iko, 2003)

Tahap-tahap *data mining* ada 6 yaitu :

1. Pembersihan data (*Data Cleaning)*

   Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian - isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut - atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data - data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi Data (*Data Integration)*

   Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa *database* atau file teks. Integrasi data dilakukan pada atribut - aribut yang mengidentifikasikan entitas - entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Transformasi Data (*Data Transformation)*

   Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

4. Proses *Mining*

   Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data

5. Evaluasi Pola (*Pattern Evaluation)*

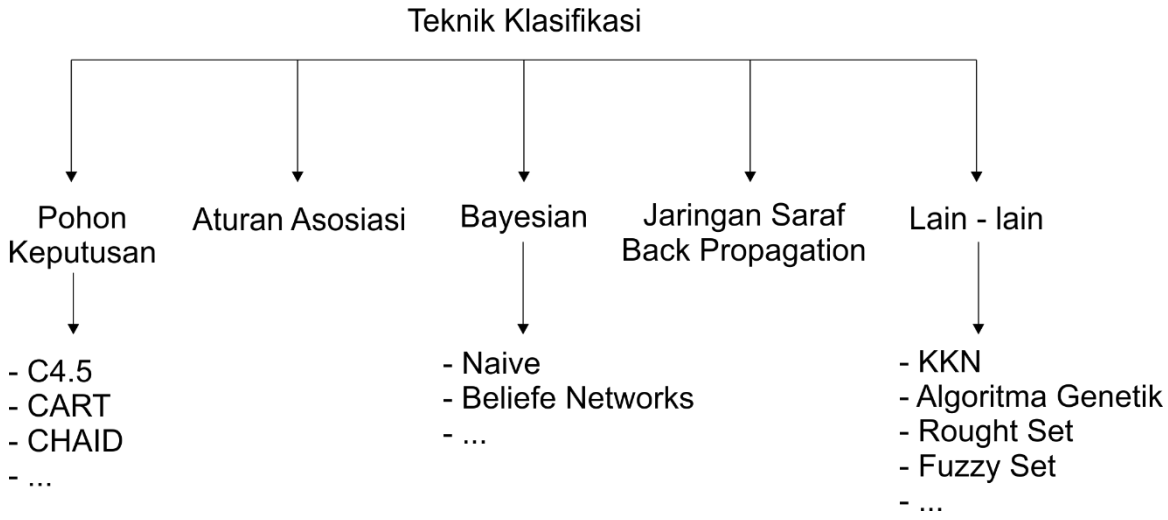
   Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

6. Presentasi Pengetahuan (*Knowledge Presentation)*

   Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining (Han, 2006)

2.3 Klasifikasi

Klasifikasi data adalah suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model Klasifikasi yang ditetapkan. Tujuan dari klasifikasi adalah untuk menemukan model dari *training set* yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya. Teknik klasifikasi terbagi menjadi beberapa teknik yang diantaranya sebagai berikut[2]:



Gambar 2.2 Teknik Klasifikasi
Sumber : (Ida, 2013)

2.4 Pohon Keputusan (*Decision Tree*)

2.4.1 Pengertian Pohon Keputusan

Decision tree adalah *flow-chart* seperti *struktur tree*, dimana tiap *internal node* menunjukkan sebuah *test* pada sebuah atribut, tiap cabang menunjukkan hasil dari *test*, dan *leaf node* menunjukkan *class-class* atau *class distribution*.[2]

2.4.2 Model Pohon Keputusan

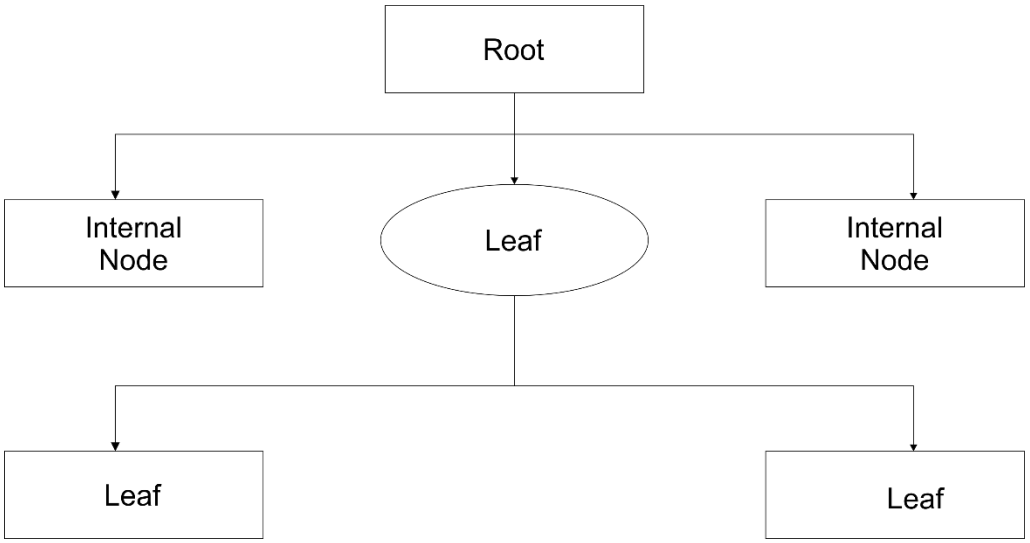
Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. *Decision tree* merupakan metode klasifikasi yang paling popular digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami. Pada *decision tree* terdapat 3 jenis *node*, yaitu:

a. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.

b. *Internal Node* , merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output minimal* dua.

c. *Leaf node* atau terminal *node* , merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

Contoh dari pohon keputusan dapat dilihat di Gambar 2.3 berikut ini :



Gambar 2.3 Model Pohon Keputusan
Sumber : (Hamidah, 2012)

2.5 Algoritma C4.5

2.5.1 Pengertian Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma *decision tree*. Algoritma ini mempunyai input berupa *training samples* dan *samples*. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field-field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data [2].

Algoritma C4.5 adalah pengembangan dari algoritma ID3. Oleh karena pengembangan tersebut algoritma C4.5 mempunyai prinsip dasar kerja yang sama dengan algoritma ID3. Hanya saja dalam algoritma C4.5 pemilihan atribut dilakukan dengan menggunakan *Gain Ratio* dengan rumus :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan :

S = Himpunan Kasus

A = Atribut

i = Jumlah Partisi Atribut

|Si| = Jumlah Kasus pada partisi ke i

|S| = Jumlah Kasus dalam S

Sedangkan penghitungan nilai *entropy* dapat dilihat pada rumus berikut

$$Entropy(S) = \sum_{i=1}^{n} - pi * \log 2\ pi \qquad (2)$$

Dengan diketahui:

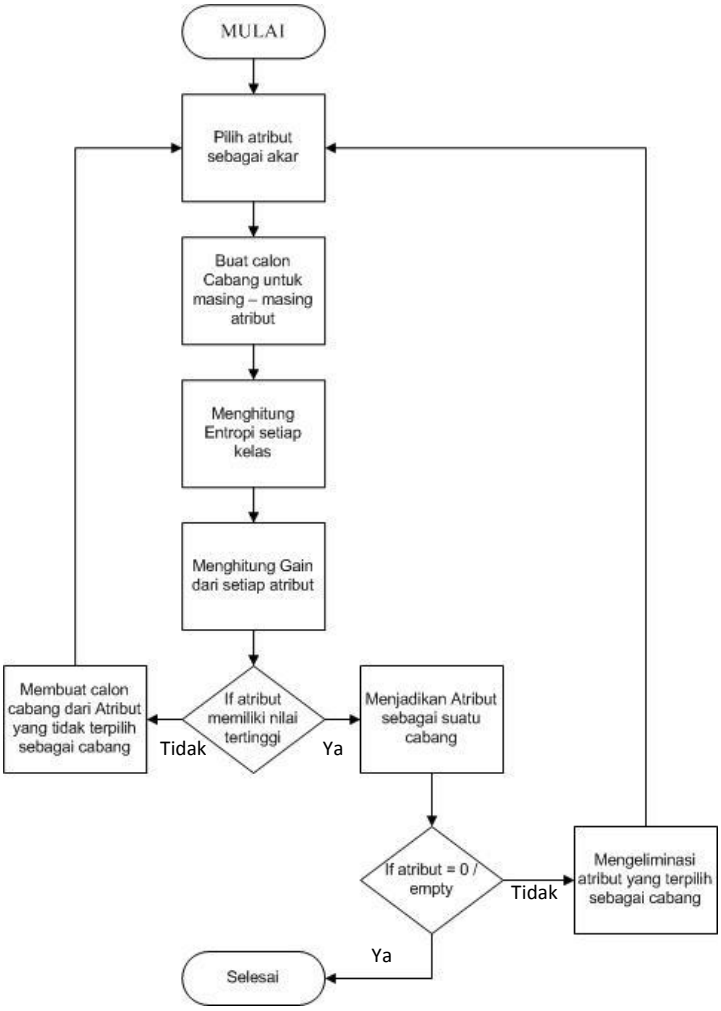
S : Himpunan Kasus

n : Jumlah partisi S

pi : Proporsi dari S_i terhadap S

Secara umum algoritma C4.5 dapat dibangun menggunakan pohon keputusan dengan cara :

a. Pilih atribut sebagai akar.
b. Buat cabang untuk masing-masing nilai.
c. Menghitung entropi kelas dan gain dari setiap atribut
d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang telah habis.



Gambar 2.4 *Flowchart* Algoritma C4.5

Sebagai contoh, untuk membentuk pohon keputusan dapat dilakukan dengan menentukan atribut sebagai akar lalu mencari nilai *Entropy* dan *Gain*. Misalkan ditampilkan 14 dataset pengujian untuk klasifikasi C4.5 sebagai berikut :

Tabel 2.1 *Date set* pengujian klasifikasi
Sumber : (Kusrini dan Luthfi, 2010)

No.	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Cloudy	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	Yes
7	Cloudy	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes
11	Sunny	Mild	Normal	TRUE	Yes
12	Cloudy	Mild	High	TRUE	Yes
13	Cloudy	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

Untuk dapat menentukan nilai-nila *Gain* dan *Entropy* dari masing – masing atribut di atas, maka terlebih dikonversikan data ke dalam bentuk tabel klasifikasi yang lebih detail sebagai berikut :

Tabel 2.2 Klasifikasi perhitungan *Gain* dan *Entropy*
Sumber : (Kustrini dan Luthfi, E.T. 2010)

Node	Kriteria	Kategori	Jumlah Kasus (S)	No (s1)	Yes (s2)	Entropy	Gain
1	Total		14	4	10	0.863	
	Outlook						0.258
		Cloudy	4	0	4	0	
		Rainy	5	1	4	0.721	
		Sunny	5	3	2	0.97	
	Temperature						
		Cool	4	0	4	0	
		Hot	4	2	2	1	
		Mild	6	2	4	0.918	
	Humidity						0.37
		High	7	4	3	0.985	
		Normal	7	0	7	0	
	Windy						0.005
		FALSE	8	2	6	0.811	
		TRUE	6	4	2	0.918	

Setelah tabel 2 dibuat, maka langkah selanjutnya adalah melakukan perhitungan dengan mencari nilai *Entropy* menggunakan persamaan (2). Setelah menentukan nilai *Entropy* langkah selanjutnya yaitu mencari nilai *Gain* menggunakana persamaan (1).

Langkah pertama yang harus dilakukan yaitu dengan mencari nilai *Entropy* total sebagai berikut :

Baris Total Kolom *Entropy* pada Tabel 2 di hitung dengan Rumus :

$$Entropy(S) = \sum_{i=1}^{n} - pi * \log_2 pi$$

Entropy (Total)

= ((-4/14_*-1,807) + ((-10/14) * (-0,385)

= 0,516 + 0,346

= 0,863

Langkah selanjutnya, mencari nilai Entropy untuk masing-masing atribut, misalnya pada Outlook (Cloudy, Rainy dan Sunny) sebagai berikut:

EntropyOutlook(Cloudy) = (-0/4 * Log2(0/4) + (-4/4) * Log2 (4/4)

= 0 + 0

= 0

EntropyOutlook(Rainy) = (-1/5) * Log2 (1/5) + (-4/5) * Log2 (4/5)

= 0,464 + 0,257

= 0,721

EntropyOutlook(Sunny) = (-3/5) * Log2 (3/5) + (-3/5) * Log2 (2/5)

= 0,442 + 0,528

= 0,970

Perhitungan selanjutnya adalah menghitung nilai *Gain* total untuk *Outlook* dengan menggunakan persamaan (1) berdasarkan nilai *Entropy* dari masing-masing atributnya sebagai berikut :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i) \qquad (1)$$

Information Gain (Total,Outlook)

= 0,863 - ((4/14 * 0) + (5/14 * 0,721) + (5/14 * 0,970))

= 0,863 - (0 + 0,240 + 0,3467) = 0,863 - 0,604

= 0,258

Perhitungan nilai *Entropy* dan *Gain* dilakukan sampai atribut terakhir sehingga dapat diperoleh nilai-nilai *Gain* dan *Entropy* tertentu. Hasil dari nilai *Gain* kemudian dibandingkan dengan nilai *Gain* lainnya untuk mencari nilai tertinggi sehingga *Gain* yang merupakan nilai tertinggi akan dijadikan acuan untuk melakukan proses perhitungan kembali pada proses pembentukan struktur pohon keputusan (Kusrini dan Luthf, 2010).

2.6 Pengertian Konsinyasi

Konsinyasi adalah hubungan kerjasama penjualan antara penyalur penjualan barang dengan pemilik barang, Dimana pemilik barang akan memasukkan barang ketoko peyalur penjualan untuk di *display* agar bisa dilihat dan dibeli oleh calon pembeli dan pemilik toko akan membayar sejumlah nilai nominal netto dari barang yang terjual dengan ketentuan yang telah disepakati. Penjualan konsinyasi sering dikenal juga titip jual.

Berikut proses pelaksanaan transaksi penjualan konsinyasi untuk toko skala besar:

1. Sebelum pemilik produk menempatkan produk ditoko, Awalnya akan dilakukan pembicaraan terlebih dahulu mengenai tempat untuk mendisplay produk pada area toko dan hal- hal lain yang dipandang perlu.
2. Pemilik barang akan memasang big wall (tidak wajib), Rak dan sarasi lainnya untuk mendukung penjualan produk ditoko.
3. Pemilik barang akan mempersiapkan SPG guna mengurus segala sesuatu yang berhubungan dengan aktivitas penjualan produk ditoko. Gaji SPG ditanggung pemilik barang yang menempatkan SPG tersebut.
4. Setelah poin 1, 2 dan 3 dilaksanakan, Kemudian pemilik barang akan mengirim barang untuk didispaly pada temapt yang sudah ditentukan untuk

dijual kepada pengunjung toko. Disini peranan SPG dan dukungan dari fihak toko sangat sangat diperlukan.

5. Setelah terjadi penjualan dalam perhitungan satu bulan, Biasanya perhitungan dalam bulan dan bukan perhitungan jumlah hari.. Misal mulai dibuka tgl 7 Oktober, Pembayaran pertama yang dilakukan atas penjualan tanggal 7 Oktober sampai dengan tanggal 31 Oktober. Dan untuk pembayaran kedua dan selanjutnya atas penjualan dalam satu bulan penuh.