



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

2.1 Internet

Menurut pendapat Supriyanto (2008) *Internet* merupakan singkatan dari *Interconnection Networking*. *Internet* berasal dari bahasa latin “*inter*” yang berarti antara. Secara kata perkata *internet* berarti jaringan antara atau penghubung, sehingga dapat disimpulkan bahwa *Internet* ialah merupakan penghubung antara suatu operasi sistem yang berbeda dengan menggunakan teknologi (telepon dan satelit) yang menggunakan protokol standar dalam berkomunikasi yaitu protokol TCP/IP (*Transmission Control/Internet Protocol*).

2.1.1 Dampak Positif dan Negatif Menggunakan Internet

Dampak positif yang diberikan *internet* sebagai berikut

- 1) *Internet* sebagai media komunikasi, yang memberikan kemudahan dalam berkomunikasi yang tidak dibatasi jarak dan ruang waktu.
- 2) Media pertukaran data, dengan menggunakan *email*, *newsgroup*, *ftp* dan *www (world wide web)* yang dapat menjadikan pengguna *internet* dapat berbagi informasi.
- 3) Media untuk mencari informasi atau data, internet memberikan kemudahan dalam mencari informasi yang akurat dan terpercaya, sehingga pengguna internet dapat mendapatkan informasi dan data yang diinginkan diseluruh dunia.
- 4) Informasi berupa pendidikan, kebudayaan dan informasi lain-lainnya.
- 5) Dapat memberikan kemudahan dalam bertransaksi.

Selain memiliki dampak positif *internet* juga memiliki dampak negatif dalam perkembangannya. Dampak negatif internet sebagai berikut:

- 1) Pornografi



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dalam mengakses internet tidak terlepas dari pornografi yang merajalela diberbagai situs di *internet*. Terlebih lagi banyak gambar-gambar dan video yang berbau pornografi dan kriminalitas yang dapat merusak moral pengguna *internet* yang melihatnya. Banyak situs penjualan yang meletakkan konten yang berbau pornografi untuk target pemasaran.

2) Penipuan

Penipuan tidak terlepas dari internet, banyak para penipuan menggunakan media internet untuk menjebak para korbannya.

3) Carding

Karena sifatnya yang *real time* (langsung), cara belanja dengan menggunakan Kartu kredit adalah cara yang paling banyak digunakan dalam dunia *internet*. Dengan sifat yang terbuka, para penjahat mampu mendeteksi adanya transaksi (yang menggunakan Kartu Kredit) *on-line* dan mencatat kode Kartu yang digunakan. Untuk selanjutnya mereka menggunakan data yang mereka dapatkan untuk kepentingan kejahatan mereka.

4) Perjudian

Perjudian tidak hanya dilakukan didunia nyata, banyak situs-situs perjudian menggunakan *internet* menjadikan penjudi tidak bermain ditempat perjudian.

2.2 Data Mining

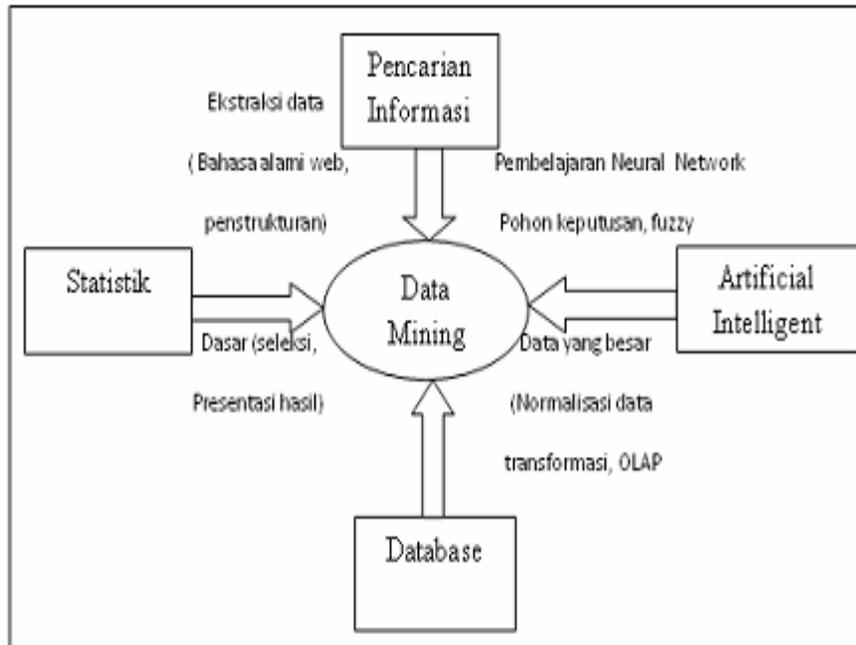
Data Mining merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam *database*. *Data Mining* juga merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database*. (Kusrini. Emha Taufiq Luthfi dikutip oleh Sari, 2013).

Hal penting yang terkait dengan *Data Mining* adalah:

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Tujuan data *mining* adalah mendapatkan hubungan atau pola yang akan mungkin memberikan indikasi yang bermanfaat.



Gambar 2.1 Bidang Ilmu Data Mining

Sumber: (Kusrini, Emha Taufiq Luthfi, 2009)

2.2.1 Fungsi Data Mining

Menurut Maclellan (2009) terdapat beberapa fungsi dari data *Mining* yang dapat digunakan, pada masalah tertentu fungsi data *Mining* dapat digunakan untuk menjawab permasalahan yang dihadapi. Berikut fungsi data *Mining* secara umum:

1. Klasifikasi (*Clasification*)

Klasifikasi sendiri memiliki fungsi untuk mengelompokkan *class* dengan kategori yang terpilih. Algoritma yang digunakan antara lain: *nearest neighbor*, pohon keputusan, *naïve bayes*, *neural networks* dan *support vector machines*.

2. *Clustering*

Clustering memiliki fungsi untuk menemukan pengelompokan attribute kedalam segmentasi-segmentasi yang sesuai similitas. Algoritma yang digunakan dalam proses *clustering* adalah *K-means*.



3. *Association*

Association berfungsi untuk mencari hubungan atribut atau *item set*, berdasarkan jumlah *item* yang muncul dari *rule association* yang ada. Algoritma pada aturan asosiasi ini adalah *Apriori*, *FP-Growth* dan *FOLD Growth*.

4. *Regretion*

Regretion memiliki kemiripan fungsi dengan klasifikasi. Adapun fungsi *regretion* adalah bertujuan untuk mencari prediksi dari suatu pola yang ada.

5. *Forecasting*

Forecasting memiliki fungsi untuk mendapatkan ramalan waktu yang didapat dengan kejadian yang *trend* diwaktu yang sebelumnya.

6. *Sequence Analysis*

Sequence Analysis memiliki fungsi untuk mencari pola urutan dari rangkaian kejadian.

7. *Deviation Analysis*

Deviation Analysis berfungsi untuk mendapatkan kejadian yang tidak sesuai dengan kejadian biasanya (Kejadian Abnormal).

2.2.2 Knowledge Discovery in Database (KDD)

Data Mining dan *Knowledge Discovery in Database* (KDD) sering kali digunakan untuk memberitahukan proses pencarian informasi yang tersembunyi didalam *database* yang besar. Pada dasarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi saling berkaitan antara satu sama lain. Dan pada salah satu tahap KDD terdapat proses *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Fayyad, 1996 dikutip oleh Kusriani, Emha Taufiq Luthi, 2009):

1. *Data Selection*

Pemilihan data dari sekumpulan data operasional yang perlu dilakukan sebelum penggalian informasi dalam KDD dimulai. Data dari hasil seleksi digunakan untuk proses *data mining*, disimpan pada suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing/Cleaning*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tahap *cleaning* ini yaitu dengan membersihkan data yang berupa duplikat, data inkonsisten. Kemudian akan dilakukan pemeriksaan seperti kesalahan dalam menulis (*tipografi*). Pada tahap ini juga dilakukan proses *enrichment*, yaitu suatu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan, seperti data atau informasi eksternal.

3. Transformation

Tahap ini dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

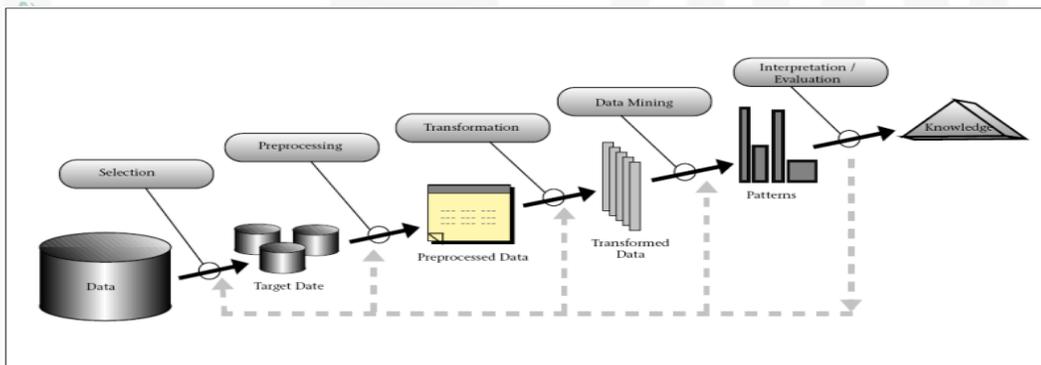
4. Data Mining

Merupakan proses mencari pola atau informasi yang sesuai dalam data yang terpilih dengan menggunakan teknik atau metode tertentu. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD.

5. Interpretation/Evaluation

Pola hasil informasi dari keluaran dalam proses *data mining* perlu dirubah dalam bentuk yang mudah dimengerti oleh orang yang bersangkutan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

Proses KDD dapat terlihat penjelasannya pada gambar 2.2 :



Gambar 2.2 Tahapan DataMining

Sumber: (Fayyad, 2006)



Hak Cipta Dilindungi Undang-Undang

2.3 Association Rules

Association rules merupakan salah satu metode didalam data *mining* untuk menentukan hubungan antar *item* dalam suatu *dataset* (sekumpulan data) yang telah ditentukan. Aturan asosiasi memberikan informasi dalam bentuk “*if...then...*” atau “*jika...maka...*” merupakan pengetahuan yang dihasilkan dari fungsi aturan Asosiasi. Istilah jika diwakili dengan *antecedent* sedangkan maka diwakili dengan istilah *consequent*.

Ada 2 parameter dalam *Asosiasi Rule* yaitu *support* (nilai penunjang) yaitu suatu nilai yang menunjukkan seberapa besar tingkat dominasi suatu *itemset* dari keseluruhan transaksi atau *dataset* dan *confidence* (nilai kepastian) yaitu suatu nilai yang menunjukkan hubungan dua *item* atau lebih secara *conditional*. Untuk mencari nilai *support 1-itemset* menggunakan persamaan 2.1 sedangkan untuk mencari nilai *2-itemset* menggunakan persamaan 2.2 di bawah ini.

$$Support(A) = \frac{JumlahTransaksiyangMengandungA}{TotalTransaksi} \times 100\% \dots \dots \dots (2.1)$$

$$Support(A, B) = \frac{JumlahTransaksiyangMengandungAdanB}{TotalTransaksi} \times 100\% \dots \dots \dots (2.2)$$

Penjelasan contoh *confidence* Misalnya “jika A maka B” bentuk demikian disebut sebagai pola hubungan. Jika pola hubungan nilai *confidence* dibawah *threshold* maka bisa diabaikan (Putri, 2013). Persamaan rumus *confidence* dijelaskan pada persamaan 2.3 di bawah ini :

$$Confidence(A \rightarrow B) = \frac{Jumlah\ Transaksi\ yang\ Mengandung\ A\ dan\ B}{Jumlah\ Transaksi\ yang\ Mengandung\ A} \times 100\% \dots \dots (2.3)$$

Salah satu algoritma yang merupakan teknik asosiasi pada data *Mining* adalah algoritma *Frequent Pattern Growth (FP-Growth)*.

2.4 Algoritma Frequent Pattern Growth (FP-Growth)

FP-Growth adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan data yang paling sering muncul (*frequent itemset*) dalam sekumpulan



- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

data (David Samuel). Struktur data yang digunakan untuk mencari *frequent Itemset* dengan algoritma *FP-Growth* adalah perluasan dari sebuah pohon *prefix*, yang biasa disebut *FP-Tree*. Dengan menggunakan *FP-Tree*, algoritma *FP-Growth* dapat langsung mengekstrak *frequent itemset* dari *FP-Tree* yang telah dibentuk dengan menggunakan prinsip *devided and conquer*.

2.4.1 Pembentukan Pohon dengan *FP-Tree*

FP-Tree merupakan struktur penyimpanan data yang dimampatkan. *FP-Tree* dibangun dengan memetakan setiap data transaksi kedalam setiap lintasan tertentu dalam *FP-Tree*. Karena dalam setiap transaksi yang dipetakan, mungkin ada 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 (Ririanti, 2014).

Pembangunan *FP-Tree* diawali dari pembentukan sebuah akar yang diberi *label null*, kemudian terdiri dari sekumpulan pohon yang beranggotakan *item-item* tertentu. Pada setiap simpul *FP-Tree* memiliki beberapa informasi, yaitu :

a. *Label item*.

Label item menjelaskan jenis *item* yang dipresentasikan oleh simpul tersebut.

b. *Support count*.

Support count mepresentasikan total lintasan transaksi yang melalui simpul tersebut.

c. *Pointer*.

Merupakan penghubung yang menghubungkan simpul-simpul dengan label *item* sama pada tiap lintasan yang ditandai dengan garis panah putus-putus.

Penjelasan pembentukan *FP-Tree* pada sebuah kasus dibawah ini dengan transaksi dinamakan TID Dapat dilihat pada tabel 2.1 dibawah ini:



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2.1 Transaksi mentah

TID	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 2.2 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



Hak Cipta Dilindungi Undang-Undang

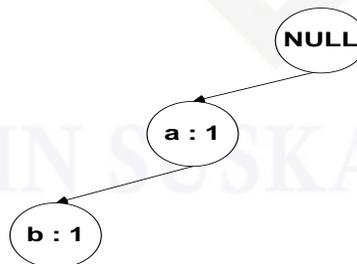
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Setelah dilakukan pemindaian pertama didapat *item* yang memiliki frekuensi di atas *support count* $\xi=20$ adalah a,b,c,d, dan e. Kelima *item* inilah yang memenuhi *support count* diatas 20% dan akan dimasukkan kedalam *FP-tree*, selebihnya (r,z,g, dan h) dapat dihilangkan karena tidak berpengaruh signifikan. Tabel berikut mendaftarkan kemunculan *item frequent* dalam setiap transaksi, diurut berdasarkan frekuensi paling tinggi.

Tabel 2.3 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}

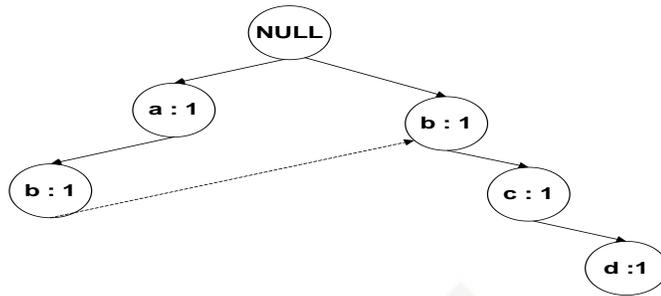
Gambar di bawah ini memberikan ilustrasi mengenai pembentukan *FP-tree* setelah pembacaan TID 1.



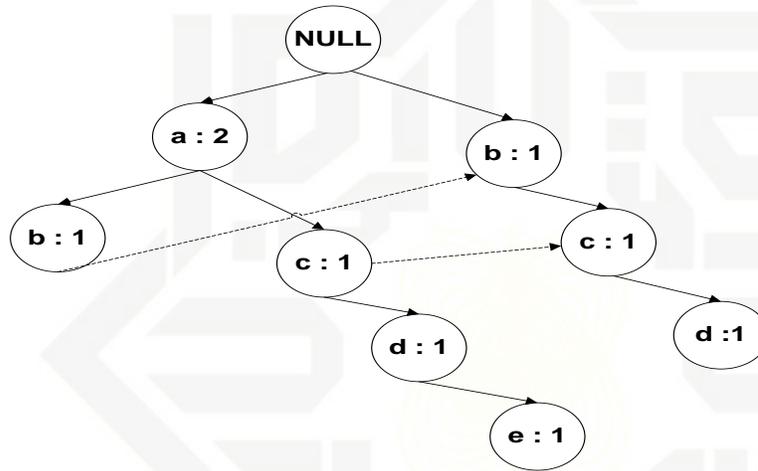
Gambar 2.3 Hasil pembentukan FP-tree setelah pembacaan TID 1

Hak Cipta Dilindungi Undang-Undang

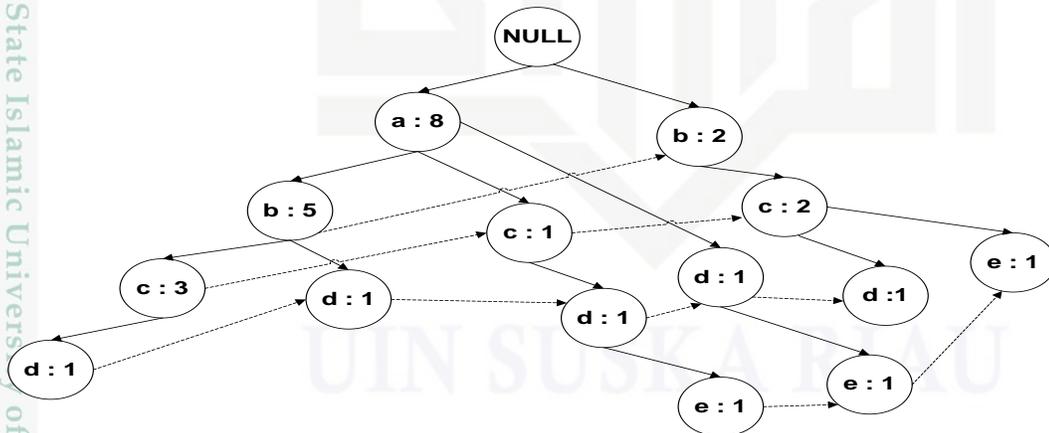
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2.4 Hasil Pembentukan *FP-Tree* setelah pembacaan TID 2



Gambar 2.5 Hasil Pembentukan *FP-Tree* setelah pembacaan TID 3



Gambar 2.6 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.

2.4.2 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 :

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* ξ 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.

Ketiga tahap tersebut merupakan langkah yang akan dilakukan untuk mendapat *frequent itemset*, yang dapat dilihat pada algoritma berikut :

Hak Cipta Dilindungi Undang-Undang

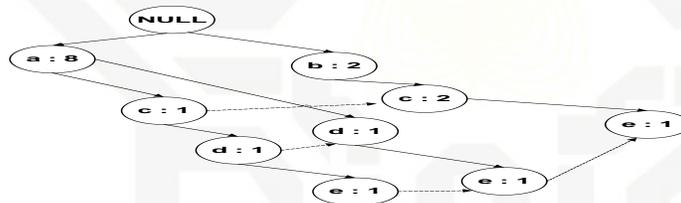
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

```

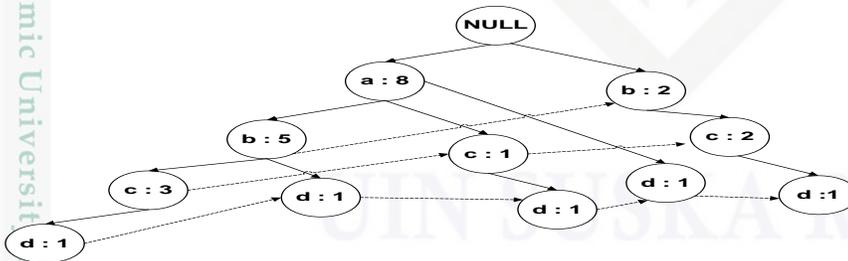
Input: FP-tree Tree
Output: Rc sekumpulan lengkap pola
       frequent
Method: FP-growth(Tree, null)
Procedure: FP-growth(Tree, α)
{
01: if Tree mengandung single path P;
02: then untuk tiap kombinasi (dinotasikan
       β) dari node-node dalam path P do
03:   bangkitkan pola β α dengan support =
       minimum support dari node-node
       dalam β;
04: else untuk tiap ai dalam header dari
       Tree do {
05:   bangkitkan pola
06:   bangun β= ai α dengan
       support = ai.support
07:   if Tree β=∅
08:   then panggil FP-growth(Tree, β) }
}
    
```

Gambar 2.7 Algoritma *FP-Growth*

Berdasarkan contoh kasus pada tahapan pembangunan *FP-tree* sebelumnya maka akan dilakukan penerapan algoritma *FP-Growth* untuk mencari *frequent itemset*. Berdasarkan kasus tersebut maka perlu menentukan upapohon lintasan yang berakhir dengan *support count* terkecil, yaitu e. Akan dicari juga upapohon secara berturut-turut untuk *support count* yang berakhir di d, c, b dan a. Proses pembentukan upapohon tersebut dapat dilihat pada gambar di bawah ini.



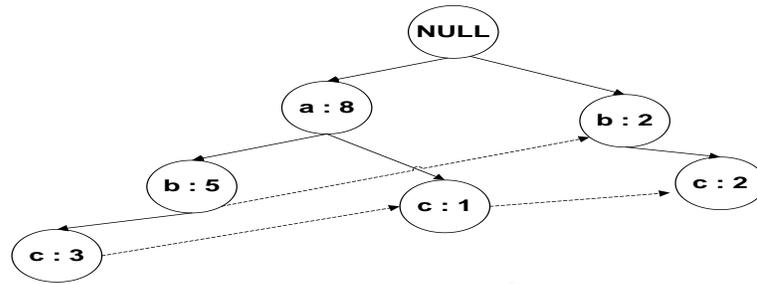
Gambar 2.8 Lintasan yang mengandung simpul e



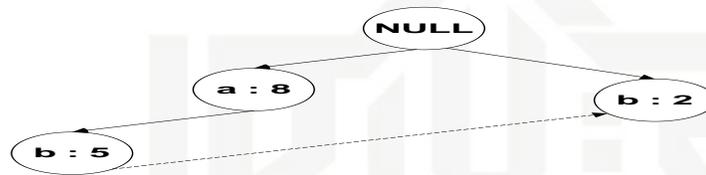
Gambar 2.9 Lintasan yang mengandung simpul d

Hak Cipta Dilindungi Undang-Undang

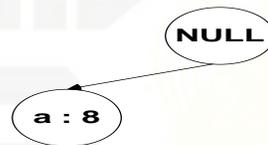
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2.10 Lintasan yang mengandung simpul c



Gambar 2.11 Lintasan yang mengandung simpul b



Gambar 2.12 Lintasan yang mengandung simpul a

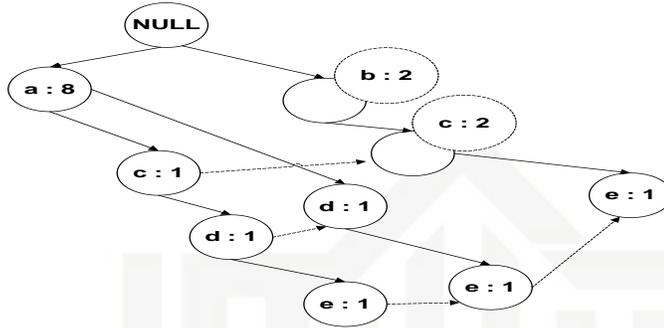
Metode *divide and conquer* digunakan oleh algoritma *FP-Growth* untuk menemukan *frequent itemset* yang berakhiran *suffix* tertentu agar dapat memecahkan *subproblem* yang lebih kecil.

Contoh, jika kita akan menemukan semua *frequent itemset* yang berakhiran e. Langkah pertama yang dilakukan adalah mengecek apakah *support count* dari e memenuhi minimum *support count* $\xi = 20\%$. Berdasarkan contoh yang ada *support count* dari e adalah 3 dan $3 \geq \xi$ maka e merupakan *item* yang *frequent*. Setelah diketahui *item* merupakan *item frequent*, maka *subproblem* selanjutnya adalah menemukan *frequent itemset* yang berakhiran de, ce, be dan ae. Dengan menggabungkan semua solusi dari *subproblem* yang ada sehingga himpunan semua *frequent itemset* yang berakhiran *item* bisa ditemukan.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Untuk lebih memperjelas penjelasan diatas, dapat dilihat pada ilustrasi dan dan langkah-langkah dibawah ini untuk menentukan *frequent itemset* yang berakhiran pada *item* di bawah ini.



Gambar 2.13 Ada lintasan yang tidak berakhiran di e, yaitu $\text{Null} \rightarrow b \rightarrow c$

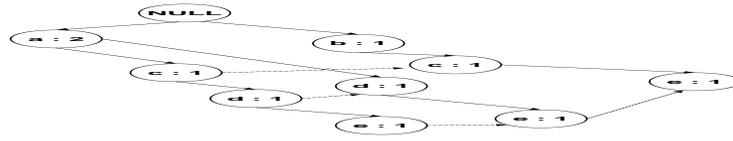
1. Bangunlah sebuah upapohon *FP-tree* yang hanya mengandung lintasan yang berakhir di e.
2. Menghitung *support count* dari *item* e kemudian dibandingkan dengan minimum *support count*. *Support count* *item* e adalah 3 maka *item* e termasuk *frequent itemset*, karena *support count* = 2.
3. Karena *item* e merupakan *item* yang *frequent*, maka perlu dipecahkan *subproblem* agar dapat menemukan *frequent itemset* yang berakhiran de, ce, be dan ae. Sebelum memecahkan *subproblem* ini, upapohon *FP-tree* tersebut harus diubah terlebih dahulu menjadi *conditional FP-tree* yang bertujuan untuk mencari *frequent itemset* yang berakhiran *item* tertentu.
4. Cara membentuk *conditional FP-tree* adalah sebagai berikut :
 - a. Buanglah setiap lintasan yang tidak mengandung *item* e. Lintasan dibagian terkanan pada contoh tidak mengandung e, yaitu $\text{Null} \rightarrow b \rightarrow c$ maka lintasan ini dapat dibuang dengan cara mengurangi *support count* dari 2 menjadi 1. Sehingga lintasan tersebut hanya mengandung transaksi {b,c,e}. Dapat dilihat pada gambar 2.14 di bawah ini.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

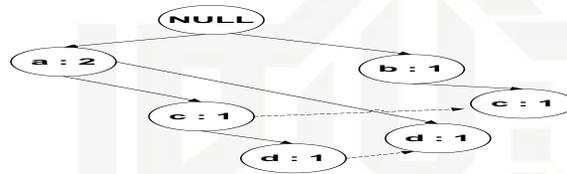
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



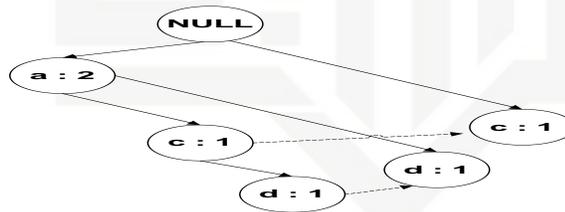
Gambar 2.14 Semua lintasan berakhir di e

- b. Setelah semua lintasan berakhir di e, maka simpul e dibuang seperti pada gambar 2.15 *Subproblem* selanjutnya yang harus dipecahkan adalah mencari lintasan *frequent itemset* yang berakhir pada de, ce, be dan ae.



Gambar 2.15 Lintasan e sudah dihapus

- c. Nilai *support count* dari b adalah 1 berarti hanya ada 1 transaksi yang mengandung b dan e maka berdasarkan prinsip *anti-monotipe heuristic*, simpul b dan lintasan yang mengandung be dapat dibuang. Hal ini dikarenakan jika *item* b tidak *frequent* maka setiap transaksi yang berakhir be juga tidak *frequent*. Sehingga terbentuklah *conditional FP-tree* untuk e, seperti pada gambar 2.16 berikut.

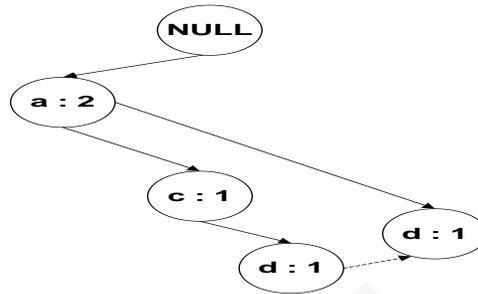


Gambar 2.16 *Conditional FP-tree* untuk e

5. Agar dapat membangun pohon lintasan *prefix* maka *FP-tree* menggunakan *Conditinal FP-tree* untuk menukan *frequent itemset* yang berakhir pada pasangan *item* de,ce dan ae.
6. Lintasan *prefix* de yang dibentuk dari *conditional FP-tree* untuk *item* e, seperti pada gambar 2. 17 berikut.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2.17 Pohon *prefix* yang berakhir di d

7. Dengan menjumlahkan *support count* dari d, yang merupakan jumlah *frequent itemset* yang berakhir di de, didapat bahwa {d,e} juga termasuk dalam *frequent itemset*
8. Selanjutnya algoritma *FP-tree* akan mengulangi langkah ketiga, sehingga didapatkan *conditional FP-tree* untuk de hanya berisi satu daun, yaitu a, dengan *support count* 2. Sehingga {a,d,e} termasuk dalam *frequent itemset*.
9. *Subproblem* berikutnya yaitu dengan menemukan *frequent itemset* yang berakhir dengan ce. Didapat {c,e} juga merupakan *frequent itemset*. Begitu juga dengan {a,e}.

Setelah memeriksa *frequent itemset* untuk beberapa akhiran (*suffix*), sehingga didapat hasil akhir seperti yang terlihat pada tabel 2.4 di bawah ini :

Tabel 2.4 Hasil *frequent itemset*.

<i>Suffix</i>	<i>Frequent Itemset</i>
E	{e}, {d,e}, {a,d,e}, {c,e}, {a,e}
D	{d}, {c,d}, {b,c,d}, {a,c,d}, {b,d}, {a,b,d}, {a,d}
C	{c}, {b,c}, {a,b,c}, {a,c},
B	{b}, {a,b}
A	{a}

Dengan metode *divide and conquer* ini, maka pada setiap langkah *rekursif*, algoritma *FP-Growth* akan membangun sebuah *conditional FP-tree* baru yang telah



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

diperbaharui nilai *support count*, dan membuang lintasan yang mengandung *item-item* yang tidak *frequent* lagi.

Setelah didapatkan *frequent itemset*, selanjutnya membuat *rule* dengan cara menghitung *confidence*-nya. Dari 19 *itemset* yang dihasilkan pada tabel 2.4, tidak semua dihitung karena *rule* yang dihasilkan adalah jika membeli barang A, maka akan membeli barang B, sehingga *itemset* yang dihitung minimal berisi dua *item*. Maka yang akan dihitung *confidence*-nya adalah 14 *subsets*, yaitu: {a,d,e}, {b,c,d}, {a,c,d}, {a,b,d}, {a,b,c}, {d,e}, {c,e}, {a,e}, {c,d}, {b,d}, {a,d}, {b,c}, {a,c} dan {a,b}.

Hanya kombinasi yang lebih besar sama dengan minimum *confidence* yang akan diambil atau *strong association rule* nya saja. Adapun rumus menghitung *confidence* adalah:

$$Confidence(A \rightarrow B) = \frac{\text{Jumlah Transaksi yang Mengandung A dan B}}{\text{Jumlah Transaksi yang Mengandung A}} \times 100\% \dots\dots(2.3)$$

Sebagai contoh akan dihitung *confidence* untuk *itemset* {a,d,e} sebagai berikut:

- $a \rightarrow d \rightarrow e = 2/4 = 50\%$
- $e \rightarrow a \rightarrow d = 2/3 = 66,7\%$
- $a \rightarrow d \rightarrow e = 2/8 = 25\%$
- $d \rightarrow e \rightarrow a = 2/2 = 100\%$
- $a \rightarrow e \rightarrow d = 2/2 = 100\%$
- $d \rightarrow a \rightarrow e = 2/5 = 40\%$
- $a \rightarrow d = 4/8 = 50\%$
- $d \rightarrow a = 4/5 = 80\%$
- $d \rightarrow e = 2/5 = 40\%$
- $e \rightarrow d = 2/3 = 66,7\%$
- $a \rightarrow e = 2/8 = 25\%$
- $e \rightarrow a = 2/3 = 66,7\%$

Jika dimisalkan *minimum confidence* adalah 75 %, *minimum confidence* ini diambil diatas 75% agar tingkat akurasinya tinggi, maka yang termasuk *strong*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

association rule adalah $d \wedge e \rightarrow a, a \wedge e \rightarrow d$ dan $d \rightarrow a$, yang artinya jika konsumen membeli *item* d dan e, maka konsumen membeli *item* a juga, jika konsumen membeli *item* a dan e, maka konsumen membeli *item* d juga dan jika konsumen membeli *item* d, maka konsumen membeli *item* a juga. Berikut tabel yang berisi hasil lengkap pola-pola atau *rules* yang dihasilkan:

Tabel 2.5 Strong Association Rule

Jika membeli	Maka akan membeli	Confidence
d dan e	A	100%
a dan e	D	100%
D	A	80%
C	B	83,3%
a dan c	B	75%

Sumber : Samuel David. Institut Teknologi bandung

2.6 Lift rasio

Untuk mengevaluasi kuat tidaknya sebuah aturan asosiasi maka bisa menggunakan *lift rasio*. *Lift rasio* merupakan perbandingan antara *confidence* sebuah aturan dengan dengan nilai *benchmark confidence*. Sedangkan *Benchmark Confidence* merupakan perbandingan anatara jumlah semua *item* yang menjadi *consequent* terhadap jumlah total transaksi (Santoso, 2007).

Rumus untuk mencari *Benchmark Confidence* adalah

$$Benchmark\ Confidence = \frac{Nc}{N} \dots\dots\dots(2.4)$$

Keterangan :

Nc = Jumlah transaksi dengan *item* yang menjadi *consequent*.

N = Jumlah transaksi basis data

Sedangkan rumus untuk mencari *lift rasio*yaitu :

$$LR = \frac{Confidence\ (A,B)}{Benchmark\ Confidence\ (A,B)} \dots\dots\dots(2.5)$$

Jika nilai *lift rasio* lebih besar dari 1 maka menunjukkan *rule* atau aturan tersebut ada manfaat dan bisa digunakan. Lebih tinggi nilai *lift rasio* maka lebih besar kekuatan asosiasinya. Dibawah ini adalah contoh nilai *lift rasio* dari kasus diatas.

**Tabel 2.6 Hasil evaluasi rules**

Jika	Maka	Confidence	Support	Nc	Bc	Lift rasio
d dan e	A	$2/2 = 100\%$ (1)	$2/10 = 20\%$	8	$8/10 = 0.8$	$1/0.8 = 1.025$
a dan e	D	$2/2 = 100\%$ (1)	$2/10 = 20\%$	5	$5/10 = 0.5$	$1/0.5 = 2$
D	A	$4/5 = 80\%$ (0.8)	$4/10 = 40\%$	8	$8/10 = 0.8$	$0.8/0.8 = 1$
C	B	$5/6 = 83,3\%$ (0.83)	$5/10 = 25\%$	7	$7/10 = 0.7$	$0.83/0.7 = 1.19$
a dan c	B	$3/4 = 75\%$ (0.75)	$3/10 = 30\%$	7	$7/10 = 0.7$	$0.75/0.7 = 1.071$

2.7 RapidMiner

Alat yang digunakan untuk pelaksanaan *FP-Growth* di penelitian ini adalah RapidMiner. RapidMiner merupakan paket *open source* berupa perangkat lunak (*Software*) Rapidminer adalah sebuah solusi untuk melakukan analisis terhadap data *Mining*, *text Mining* dan analisis prediksi.

2.8 Kajian Pustaka

Pada penelitian ini belum ditemukan dengan pasti meneliti yang berkaitan dengan mencari hubungan pekerjaan orangtua dan pola internet anak menggunakan algoritma *FP-Growth* itu sendiri, namun penulis mengkaji dari beberapa penelitian yang mendekati dengan judul yang akan diteliti oleh penulis yaitu dengan judul penelitian “Penerapan metode *Frequent Pattern-Growth* (FP-Growth) untuk mencari hubungan pekerjaan orang tua dengan pola internet anak SD di Pekanbaru”.

Kemudian penelitian yang berkaitan dengan metode yang digunakan penulis yaitu penelitian yang dilakukan oleh Abdullah Saad Almalaise Alghamdi tahun 2011 dengan judul “*Efficient Implementation of FP Growth Algorithm-Data Mining on Medical Data*”. Tujuan dari penelitian ini adalah memberikan informasi yang menghubungkan beberapa atribut dari data medis berupa data penyakit dan data pasien yang membantu dokter dan pasien dalam pencegahan dan memberikan obat-obat yang sesuai dengan penyakit pasien. Penelitian ini menggunakan tahapan data



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

mining yaitu *preprocessing* dengan tujuan untuk mengekstrak pola yang tidak biasa dari data. Sebagai contoh, serangan jantung orangtua / angina sebelum 50 usia tergantung pada dimana ayah lahir, jika orang tua memiliki asma sebelum setiap usia, dan jika orang tua memiliki diabetes pada setiap usia dengan nilai keyakinan 0,952.

Penelitian yang dilakukan oleh Sidhu, Shivam, dkk pada tahun 2014 yang berjudul “*FP Growth Algorithm Implementation*”. Pada penelitian ini didapatkan output berupa 1 aturan kombinasi yang memenuhi *minimum support* = 20% dan *minimum confidence* = 80% yaitu jika perempuan dengan pendidikan terakhir adalah SMA maka tidak bisa berbicara dengan bahasa lain selain bahasa inggris. Adapun tujuan dari penelitian ini adalah mencari hubungan antara jenis kelamin, status pernikahan dan pendidikan terakhir terhadap kemampuan berbicara selain bahasa inggris sehingga proses penerapan *algoritma FP-growth* pada penelitian ini didapat untuk mendapatkan aturan dari kombinasi yang memberikan nilai yang memenuhi *minimum confidence*.