

BAB 2

TINJAUAN PUSTAKA

1.1 Profil Koperasi Waserda Bravo

Koperasi Plasma Pratama Mandiri (KPPM) adalah Koperasi Primer yang bergerak dibidang usaha dan simpan pinjam. Koperasi Plasma Pratama Mandiri terletak di Desa Bumi Pratama Mandira Kecamatan Sungai Menang Kabupaten Ogan Komering Ilir Propinsi Sumatera Selatan. Jumlah Anggota Koperasi Plasma Pratama Mandiri sebanyak 2970 orang. Koperasi ini memiliki dua cabang toko yang menjual berbagai macam kebutuhan bahan pokok yang disebut Waserda Alpha dan Waserda Bravo. Selain itu, Koperasi Plasma Pratama Mandiri (KPPM) melakukan kerjasama dengan berbagai pihak seperti Bank BRI, Honda, dan JNE Express. Sehingga, selain melayani kebutuhan bahan pokok Waserda Alpha dan Bravo juga melayani penarikan uang tunai, transfer tunai , transfer antar bank dll melalui Mini ATM BRILink, melayani jual beli kendaraan roda dua yaitu sepeda motor, dan melayani untuk pengiriman dan penerimaan barang seperti paket. Hal ini dilakukan oleh Koperasi untuk memudahkan masyarakat desa dalam memenuhi kebutuhan hidupnya.

1.2 Data Mining

Data mining adalah proses menemukan pola dan pengetahuan yang menarik dari sejumlah besar data. Sumber data dapat mencakup basis data, gudang data, web repository informasi lainnya, atau data yang dimasukkan ke dalam sistem secara dinamis (Han, Kamber, & Pei, 2012). Data mining adalah pencarian dan

teknik analisa data yang besar untuk menemukan pola dan aturan yang berarti
(Barry,



Michael, & Gordon, 2004). Data Mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis (Hermawati, 2013). Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining adalah 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 besar (Nursikuwagus & Hartono, 2016). Data mining dapat dibagi menjadi beberapa tahap proses. Tahap-Tahap Data Mining adalah sebagai berikut (Han et al., 2012):

1. Pembersihan Data (*data cleaning*) adalah untuk menghilangkan noise dan data yang tidak konsisten.
2. Integrasi Data (*data integration*) adalah dimana banyak penggabungan sumber data
3. Seleksi Data (*data selection*) adalah dimana data yang relevan yang akan diambil dari database untuk di analisis
4. Transformasi data (*data transformation*) adalah dimana data di transformasi dan konsolidasi ke dalam bentuk yang sesuai, dengan melakukan ringkasan atau agregasi operasi.
5. Proses Mining adalah tahap dimana proses yang esensial yang diterapkan untuk mengekstrak pola data
6. Evaluasi pola (*pattern evaluation*) adalah untuk mengidentifikasi pola-pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan langkah-langkah yang menarik.

7. Presentasi Pengetahuan (*Knowledge Presentation*) adalah dimana teknik visualisasi dan representasi pengetahuan digunakan untuk menyajikan pengetahuan yang ditambang kepada pengguna.

Adapun tujuan dari data mining adalah (Baskoro,2010):

1. Explanatory digunakan untuk menjelaskan beberapa kegiatan observasi.
2. Confirmatory yaitu digunakan untuk mengkonfirmasi suatu hipotesis yang telah ada.
3. Exploratory yaitu untuk menganalisa data baru dalam suatu relasi yang janggal.

Sebelum melakukan proses data mining, baiknya mengetahui terlebih dahulu apa yang bisa dilakukan oleh data mining, agar apa yang dilakukan nantinya memang sesuai dengan apa yang dibutuhkan serta menghasilkan sesuatu yang sebelumnya tidak diketahui dan bersifat baru serta bermanfaat bagi penggunanya sendiri (Prasetyo, 2012).

Data mining dapat dikelompokkan menjadi beberapa bagian berdasarkan tugas yang bisa dilakukannya yaitu: (Taufiq et al., 2009)

1. Deskripsi yaitu dimana para penulis mencoba untuk mendeskripsikan pola dan trend yang ada dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan adanya penjelasan untuk suatu pola atau kecenderungan. Contohnya yaitu pada petugas pengumpulan suara dimana tidak dapat menentukan fakta atau keterangan yang jelas siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Dalam hal ini deskripsi dari sebuah pola dan kecenderungan sering

- memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan (Sari, 2018).
2. Estimasi yaitu Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih kearah numerik dari pada kategori. Metode data mining yang terdapat di dalam estimasi yaitu, (Linear Regression, Neural Network, Support Vector). Contohnya estimasi kumulatif mahasiswa pada program pascasarjana dengan melihat dari nilai indeks prestasi (IPK) mahasiswa tersebut pada saat mengikuti program sarjana (Ukiarwan, 2017)
 3. Prediksi yaitu yang memiliki kemiripan dengan estimasi dan klasifikasi. Akan tetapi, hasilnya menunjukkan sesuatu yang belum terjadi (masa depan). Metode data mining yang terdapat di dalam prediksi yaitu, (Linear Regression, Neural Network, Support Vector). Misalnya ingin mengetahui prediksi harga beras tiga bulan yang akan datang (Sari, 2018).
 4. Klasifikasi yaitu dimana variabel tujuan bersifat kategorik. Metode data mining yang terdapat di dalam klasifikasi yaitu, (Naive Bayes, K-Nearest, C4.5, ID3, CART, Linear Discriminant Analysis, Logistic Regression). Contoh klasifikasi dalam bidang bisnis yaitu menentukan apakah suatu transaksi kartu kredit adalah transaksi yang benar atau salah (Ukiarwan, 2017).
 5. Clustering yaitu bentuk pengelompokkan data mining yang lebih kearah record, pengamatan, atau kasus dala kelas yang memiliki suatu kemiripan. Clustering lebih kearah pengelompokan record, pengamatan, atau kasus Metode data mining yang terdapat di dalam clustering yaitu, (K-Means, K-Medoids, Self-Organizing Map (SOM), Fuzzy C-Means).
 6. Asosiasi yaitu untuk mengidentifikasi hubungan antara berbagai peristiwa yang terjadi dalam satu waktu. Metode data mining yang terdapat di dalam

asosiasi yaitu, (FP-Growth, A Priori, Coefficient of Correlation, Chi Square). Contoh asosiasi dalam bidang bisnis yaitu untuk menemukan barang apa saja yang sering di beli secara bersamaan, dan barang apa saja yang tidak pernah atau jarang sekali dibeli oleh konsumen.

1.3 Aturan Asosiasi (*Association Rule*)

Analisis Asosiasi atau *Association Rules* adalah suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk support dan confidence dalam suatu database. *Association Rules* adalah suatu proses untuk mencari hubungan antar item dalam suatu data sel yang telah ditentukan (Han et al., 2012). Association rule merupakan salah satu teknik data mining yang digunakan dalam Market Basket Analysis mengasosiasikan data yang satu dengan data yang lain (Bakri, Halim, & Astuti, 2018). Association rule adalah metode yang bertujuan mencari pola yang sering muncul di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa item sehingga metode ini akan mendukung sistem rekomendasi peminjaman buku yang akan di pinjam oleh pengguna melalui penemuan pola antar item dalam transaksi-transaksi yang terjadi di Bapersip (Zhang, Zhang, & Yang, 2013).

Penerapan data mining dengan aturan asosiasi bertujuan menemukan informasi item-item yang saling berhubungan dalam bentuk aturan/rule. Aturan asosiasi adalah teknik data mining untuk menemukan aturan asosiasi antar suatu kombinasi item (Ardani & Fitriana, 2016). Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang mendasari dari berbagai teknik data mining yang lain. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis

pola frekuensi tinggi (frequent pattern mining), banyak menarik perhatian peneliti untuk menghasilkan algoritma yang efisien. Dengan menggunakan dua parameter dapat diketahui seberapa penting aturan asosiasi yang berlaku, nilai support (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan nilai kepercayaan (Wahdi, 2018).

Dalam menentukan suatu *association rule*, terdapat suatu ukuran yang menarik (ukuran kepercayaan) yang diperoleh dari hasil pengolahan data dengan menggunakan perhitungan tertentu (Ristianingrum & Sulastri, 2017). Pada umumnya ada dua ukuran, yaitu:

1. Support yaitu ukuran untuk menunjukkan seberapa besar ukuran tingkat dominasi suatu item dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu item tersebut layak untuk dicari confidence-nya. Contohnya dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi suatu item yang menunjukkan bahwa item A dan item B dibeli bersamaan.
2. Confidence: menunjukkan ukuran suatu hubungan antara 2 item berbeda secara conditional. Misalnya, menghitung kemungkinan besar seberapa sering item B dibeli oleh pelanggan jika pelanggan tersebut membeli sebuah item A

Nantinya kedua ukuran ini bisa berguna dalam menentukan kekuatan suatu pola yang akan dihasilkan, dengan cara membandingkan pola tersebut dengan nilai minimum kedua parameter tersebut yang sudah ditentukan oleh pengguna. Jika suatu pola sudah memenuhi kedua nilai minimum parameter yang sudah

ditentukan sebelumnya, maka pola tersebut dapat disebut sebagai aturan yang menarik atau aturan yang akurat.

1.4 Tahapan Association Rule

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap (Taufiq et al., 2009) yaitu:

1. Analisa pola frekuensi tinggi yaitu mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database. Nilai

support tersebut dapat diperoleh dengan rumus:

$$\text{Support}(A) = \frac{\text{Jumlah Transaksi Mengandung } A}{\text{Total Transaksi}} \quad 2.1 \text{ i}$$

Sedangkan nilai support dari 2 item diperoleh dengan rumus:

$$\text{Support}(A, B) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi}} \quad 2.2 \text{ i}$$

2. Pembentukan aturan asosiatif yaitu setelah semua pola memiliki frekuensi yang tinggi ditemukan, maka selanjutnya mencari aturan asosiasi yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiatif A U B. Dimana nilai confidence dari aturan A U B dapat diperoleh dengan rumus sebagai berikut:

$$\text{Confidence } P(B|A) = \frac{\sum \text{Transaksi Mengandung } A \text{ dan } B}{\sum \text{Transaksi Mengandung } A} \quad 2.3 \text{ i}$$

Rumus diatas menjelaskan bahwa nilai confidence diperoleh dengan cara membagi jumlah transaksi yang mengandung item A dan item B dengan kata lain item pertama bersamaan dengan item yang kedua atau lebih dengan jumlah transaksi yang mengandung item A.

1.5 Algoritma *FP-Growth*

Frequent pattern growth algorithm atau yang biasa disebut dengan FP-Growth Algorithm adalah suatu algoritma yang memperkecil ukuran data set yang merepresentasikan frequent item ke dalam frequent pattern tree (FPTree) (Han et al., 2012). Algoritma FP-Growth merupakan pengembangan dari algoritma Apriori. Algoritma Frequent Pattern Growth adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (frequent itemset) dalam sebuah kumpulan data (Samuel, 2018). Karakteristik algoritma FP-Growth adalah struktur data yang menggunakan tree atau disebut dengan Frequent Pattern Tree (FPTree). Dengan menggunakan FPTree, algoritma FP-growth dapat langsung mengekstrak frequent itemset dari FP-Tree (Setyawan, 2018). Algoritma FP-Growth menggunakan konsep pembangunan tree, yang biasa disebut FP-Tree, dalam pencarian frequent itemsets bukan menggunakan generate candidate seperti yang dilakukan pada algoritma Apriori. Dengan menggunakan konsep tersebut, algoritma FP-Growth menjadi lebih cepat daripada algoritma Apriori (Erwin, 2009). Metode FP-Growth sendiri terdiri dari tiga tahapan utama (Fitriyani, 2015) yaitu:

1. Tahap pembangkitan conditional pattern base merupakan sub-database yang berisi prefix path (pola awalan) dan suffix pattern (pola akhiran). Pembangkitan Conditional Pattern Base didapatkan melalui FP-Tree yang telah dibangun sebelumnya.
2. Tahap pembangkitan conditional FP-Tree, dimana pada tahap ini nilai frekuensi dari setiap item untuk Conditional Pattern Base dijumlahkan lalu setiap item yang memiliki jumlah nilai frekuensi lebih besar sama

dengan nilai minimum support maka akan dibangkitkan dengan Conditional FP-Tree.

3. Tahap pencarian frequent itemset. merupakan lintasan tunggal (single path), kemudian didapatkan frequent itemset dengan melakukan kombinasi item untuk Conditional FP-Tree. Jika bukan lintasan tunggal, maka dilakukan pembangkitan FP-Growth secara rekursif (fungsi yang memanggil dirinya sendiri secara langsung ataupun tidak langsung).

1.6 Pembangunan FP-Tree

Menurut (Tan, Michael, & Vipin., 2004) pembangunan FP-Tree ini dilakukan dengan cara memetakan setiap data transaksi kedalam setiap lintasan tertentu dalam FP-Tree. Kemungkinan besar dalam pemetaan item dari setiap transaksi memiliki item yang sama, sehingga lintasan yang dibuat dapat dilewati secara berulang kali atau tumpang tindih. Dan semakin banyak data transaksi item yang sama, maka proses kompresi dengan struktur data FP-Tree semakin efektif.

Kelebihan dari FP-tree adalah hanya memerlukan dua kali pemindaian data transaksi yang terbukti sangat efisien. Support adalah penghitung frekuensi kemunculan transaksi yang mengandung suatu pola. Suatu pola itu dapat dikatakan sering muncul atau *frequent pattern* apabila support dari pola tersebut tidak kurang dari konstanta atau batas ambang minimum support. Permasalahan dalam mencari pola *frequent* dengan menggunakan batas ambang minimum support count ini yang akan dicoba untuk dipecahkan oleh FP-Growth dengan bantuan struktur FP-Tree.

Adapun FP- tree yaitu sebuah pohon dengan definisi sebagai berikut:

1. FP-Tree dibentuk oleh sebuah akar yang diberi label null, sekumpulan upapohon yang beranggotakan item-item tertentu, dan sebuah tabel frequent header.
2. 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 antarlintasan, ditandai dengan garis panah putus-putus.



Contoh 1. Misal tabel data transaksi dengan minimal *support count* = 2

Table 2.1 Contoh Tabel 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 sebagai berikut:

Table 2.2.Frekuensi Kemunculan tiap item

Ite m	Frekuen si
a	8
b	7
c	6
d	5
e	3
f	1
r	1
z	1
g	1
h	1

Dari tabel hasil frekuensi kemunculan tiap item diatas, dapat kita lihat bahwa dengan nilai *support count* = 2 item yang akan berpengaruh dan dimasukkan kedalam FP-Tree adalah item a,b,c,d, dan e. Sedangkan item r,z,g, dan h dapat dibuang karena tidak berpengaruh secara signifikan. Berikut adalah tabel yang mendata kemunculan item yang *frequent* dalam setiap transaksinya dan diurutkan berdasarkan frekuensi yang terbesar.

Table 2.3 Data Transaksi

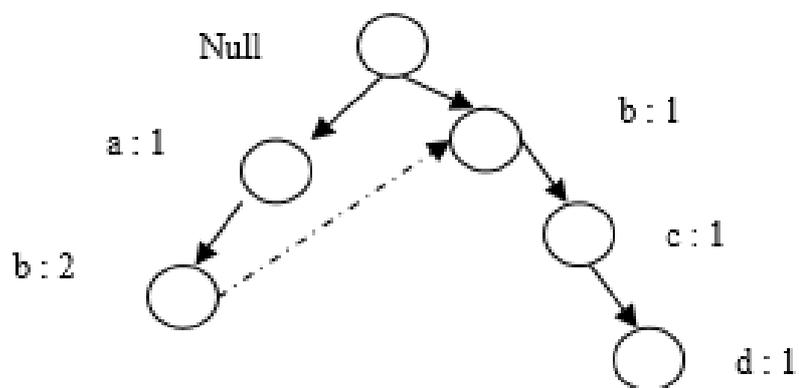
TI D	Item
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}

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

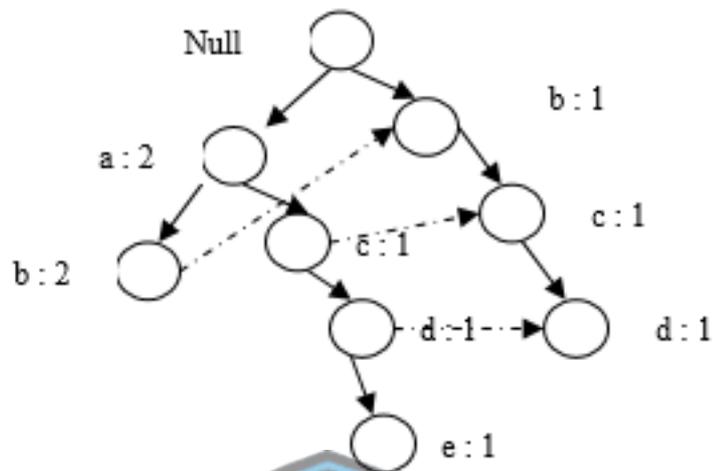
Gambar dibawah ini adalah ilustrasi mengenai pembentukan FP-Tree setelah pembacaan TID 1.



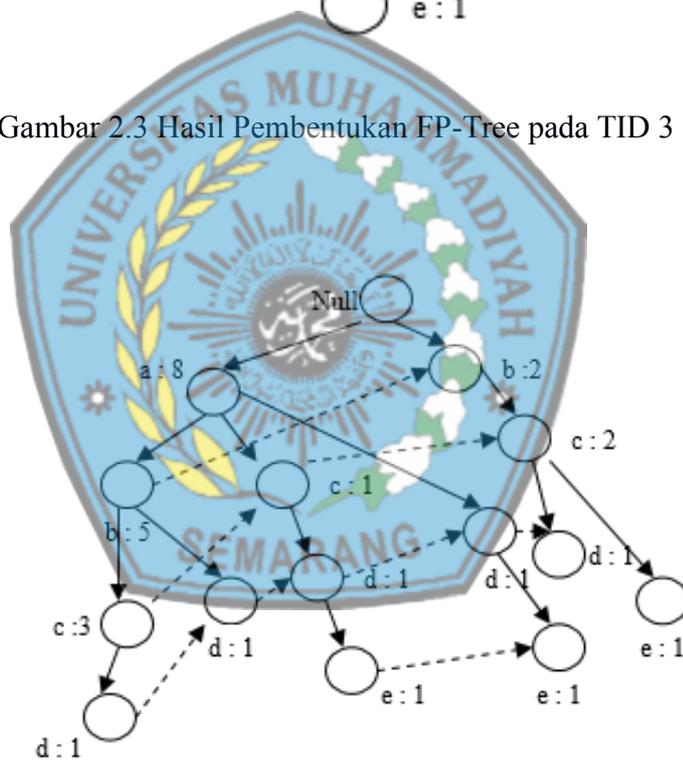
Gambar 2.1 Hasil Pembentukan FP-Tree pada TID 1.



Gambar 2.2 Hasil Pembentukan FP-Tree pada TID 2



Gambar 2.3 Hasil Pembentukan FP-Tree pada TID 3



Gambar 2.4 Hasil Pembentukan FP-Tree pada TID 10.

Gambar diatas merupakan gambar terbentuknya FP-Tree dalam setiap TID yang dibaca. Setiap simpul pada FP-Tree ini mengandung nama sebuah item dan counter support yang berfungsi untuk menghitung frekuensi kemunculan item tersebut dalam setiap lintasan transaksinya.

FP-tree yang merepresentasikan data transaksi pada tabel 2.1 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 2.1.
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 1
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.