

Pencarian Pola Asosiasi Data Nasabah Bank Menggunakan Algoritma Apriori

Hardian Oktavianto¹⁾

¹⁾Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Jl. Karimata No. 49 Jember Kode Pos 68121
Email : ¹⁾hardian@unmuhjember.ac.id

ABSTRAK

Data Mining merupakan salah satu teknik untuk menemukan, mencari, atau menggali informasi atau pengetahuan baru yang kemudian menghasilkan informasi yang berguna. Fungsi *data mining* yang sering digunakan adalah untuk klasifikasi, klasterisasi, estimasi, prediksi, serta penemuan pola asosiasi (*association rule mining*). *Association rule mining* merupakan salah satu teknik dalam *data mining* yang berguna untuk menemukan pola asosiasi tersembunyi dalam suatu basis data. Pola asosiasi yang ditemukan nantinya berupa *rule-rule* asosiasi antara *itemset* dengan masing-masing nilai bobot asosiasinya. Pada penelitian ini, akan dilakukan pencarian pola asosiasi terhadap data nasabah di suatu bank, dengan melibatkan faktor umur, jenis kelamin, domisili, pendapatan, status pernikahan, jumlah anak, kepemilikan mobil, kepemilikan rekening tabungan, kepemilikan rekening giro, kepemilikan hipotek, dan status pembelian *Personal Equity Plan*. Hasil dari penelitian ini adalah terbentuknya pola bahwa seorang nasabah akan memiliki kemungkinan 100% untuk ikut dalam program *Personal Equity Plan* jika berumur antara 35 sampai 51 tahun dan mempunyai 1 orang anak serta mempunyai rekening giro. Sebaliknya, seorang nasabah kemungkinan 100% untuk tidak ikut dalam program *Personal Equity Plan* jika berpenghasilan antara US\$ 24.387 sampai US\$ 43.758, dengan status menikah, dan belum mempunyai anak, serta memiliki rekening giro, dan tidak memiliki hipotek, selain itu seorang nasabah juga memiliki kemungkinan 100% tidak ikut dalam program *Personal Equity Plan* jika berumur mulai 52 tahun keatas, dengan status menikah, dan belum memiliki anak, serta mempunyai rekening tabungan dan rekening giro.

Kata Kunci : *Association Rule, Apriori, Personal Equity Plan, Nasabah*

1. PENDAHULUAN

Data Mining merupakan salah satu teknik untuk menemukan, mencari, atau menggali informasi atau pengetahuan baru dari sekumpulan data yang sangat besar, dengan integrasi atau penggabungan dengan disiplin ilmu lain seperti statistika, kecerdasan buatan, serta machine learning, menjadikan data mining sebagai salah satu alat bantu untuk menganalisa data yang kemudian menghasilkan informasi yang berguna. Fungsi data mining yang sering digunakan adalah untuk klasifikasi, klasterisasi, estimasi, prediksi, serta penemuan pola asosiasi (*association rule mining*).

Association rule mining merupakan salah satu teknik dalam data mining yang berguna untuk menemukan pola asosiasi tersembunyi dalam suatu basis data, pola yang dimaksud disini adalah keterkaitan atau korelasi antara tiap-tiap item yang berbeda pada setiap record di dalam basis data. Pola asosiasi yang ditemukan nantinya berupa *rule-rule* asosiasi antara *itemset* dengan masing-masing nilai bobot asosiasinya, *rule* yang terbentuk biasa dinotasikan dengan $X \rightarrow Y$ dimana X dan Y disini adalah *itemset*, bobot asosiasi disini berupa nilai *support* yang menjelaskan berapa kali sebuah *itemset* tercatat atau muncul dari sejumlah

dataset dan nilai *confidence* yang menjelaskan seberapa kuat hubungan diantara itemset X dan Y.

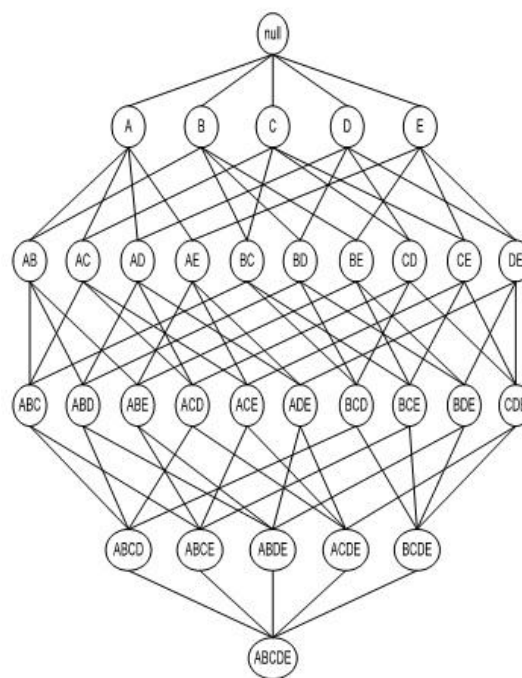
Penelitian yang mengangkat topik tentang *association rule mining* telah banyak dilakukan, diantaranya adalah penelitian yang dilakukan oleh Himigiri dan Roy pada tahun 2011 yang berjudul *Effective Data Mining Association Rules for Heart Disease Prediction System*, penelitian tersebut bertujuan untuk menemukan pola asosiasi yang memperkirakan kondisi pembuluh darah yang normal atau terserang penyakit berdasarkan faktor resiko serta faktor kesehatan lainnya. Penelitian lainnya adalah yang dilakukan oleh Srinivas, Rao, dan Govardhan pada tahun 2012 dengan judul *Mining Association Rules from Large Datasets Towards Disease Prediction*, tujuan dari penelitian tersebut adalah menemukan pola asosiasi diantara penderita penyakit jantung di daerah tambang batu bara dengan memperhatikan faktor umur dengan kondisi lingkungan.

Pada penelitian ini, akan dilakukan pencarian pola asosiasi terhadap data nasabah di suatu bank, dengan melibatkan faktor umur, jenis kelamin, domisili, pendapatan, status pernikahan, jumlah anak, kepemilikan mobil, kepemilikan rekening tabungan, kepemilikan rekening giro, kepemilikan hipotek, dan status pembelian *Personal Equity Plan*. Dataset yang digunakan adalah diambil secara online dari laman <http://facweb.cs.depaul.edu/>. Tujuan dari penelitian ini adalah menemukan pola keterkaitan antara 11 faktor yang telah disebutkan sebelumnya dengan menerapkan algoritma apriori.

2. TINJAUAN PUSTAKA

Tahap penting dari *association rule mining* adalah pencarian *frequent itemset*

dan pembentukan *association rule* (yang selanjutnya disebut *rule*). *Frequent itemset* dicari dengan menggunakan batasan nilai *minimum support* (*min_supp*), sedangkan *rule - rule* dibentuk berdasarkan nilai *minimum support* (*min_supp*) dan *minimum confidence* (*min_conf*). *Itemset* dibentuk dari masing masing *item* yang terdapat di basis data, dimulai dengan pencarian 1-*itemset*, kemudian mengkombinasikan 1-*itemset* dengan 1-*itemset* yang lainnya untuk mendapatkan 2-*itemset*, dan seterusnya, sampai dengan *k-itemset* dimana *k* adalah jumlah *item* yang berbeda pada suatu dataset, sebagai ilustrasi penjabaran, pembentukan *itemset-itemset* ditunjukkan oleh Gambar 1.



Gambar 1. Diagram Lattice *Itemset*

Gambar 1 diatas menunjukkan bahwa dataset memuat 5 buah *item* yang berbeda, yaitu A, B, C, D, dan E, serta menjelaskan *itemset-itemset* apa saja yang akan terbentuk dengan *itemset* terakhir yang terbentuk berupa 5-*itemset*, yang mana sesuai dengan total jumlah *item* berbeda.

Pembentukan *itemset* tidak memperhatikan jumlah transaksi yang ada melainkan hanya memperhatikan *item-item* yang berbeda, setelah semua *itemset* ditemukan, maka dilakukan pencarian *itemset* mana yang termasuk *frequent itemset*, dimana *frequent itemset* adalah *itemset* yang memenuhi nilai *support* lebih besar atau sama dengan batasan *minimum support (min_supp)* yang telah ditentukan sebelumnya. Hasil seleksi *itemset* inilah yang akan menjadi *frequent k-itemset*, dengan *k* adalah jumlah *item*.

Aturan asosiasi (*rule*) akan dibentuk dari tiap tiap *frequent itemset* dengan dimulai dari *frequent 2-itemset* sampai dengan *frequent k-itemset*, kemudian akan diseleksi dengan menggunakan *minimum support (min_supp)* dan *minimum confidence (min_conf)*, hasil seleksi ini kemudian akan menjadi hasil akhir dari *association rule mining*.

Hasil akhir dari *association rule mining* berupa sejumlah *rule* yang dinotasikan dengan $X \rightarrow Y$ dan atau sebaliknya $Y \rightarrow X$, masing-masing memiliki nilai *support* (*s*) dan *confidence* (*c*), dimana *X* dan *Y* disini adalah *frequent itemset*, nilai *support* menjelaskan berapa kali sebuah *itemset* tercatat atau muncul dari sejumlah dataset, sedangkan nilai *confidence* menjelaskan seberapa kuat hubungan diantara *itemset* *X* dan *Y*.

$$\text{support (s)} (X \rightarrow Y) : \frac{\text{jumlah } h(X \cup Y)}{\text{total transaksi}}$$

$$\text{confidence (c)} (X \rightarrow Y) : \frac{\text{jumlah } h(X \cup Y)}{\text{jumlah } h(X)}$$

Untuk memperjelas bagaimana hasil akhir dari *association rule mining*, bisa dilihat pada contoh berikut. Misalkan terdapat transaksi seperti dalam Tabel 1.

Tabel 1. Contoh Transaksi

IDTransaksi	ItemDibeli
1	A, B, C
2	B, C
3	A, B
4	C

Maka semua *itemset* yang mungkin adalah $\{\{A\}, \{B\}, \{C\}, \{AB\}, \{AC\}, \{BC\}, \{ABC\}\}$, dan jika diinisialisasi nilai $\text{min_supp} = 50\%$ maka *frequent itemset* yang didapat adalah $\{\{AB\}, \{BC\}\}$. Berdasarkan *frequent itemset* yang ada maka *rule* yang terbentuk adalah $A \rightarrow B$, $B \rightarrow A$, $B \rightarrow C$, dan $C \rightarrow B$. Jika ingin dicari *rule - rule* mana saja yang memenuhi nilai *minimum support* 50% dan *minimum confidence* 75% maka kita hanya mendapatkan *rule* $A \rightarrow B$ (50%, 100%), yang berarti sebanyak 50% dari total transaksi *item* *A* dan *B* dibeli secara bersamaan, dengan kemungkinan sebesar 100% *item* *B* dibeli jika *A* dibeli, sedangkan untuk ketiga *rule* yang lain diabaikan karena tidak memenuhi nilai *minimum confidence*.

Algoritma apriori pertama kali dikenalkan oleh R. Agrawal dan R. Srikant pada tahun 1994, algoritma ini menggunakan prinsip apriori dalam pencarian *frequent itemset*, yaitu semua *subset* yang tidak kosong dari sebuah *frequent itemset* pasti juga akan merupakan *frequent*, begitu juga dengan semua *superset* yang tidak kosong dari sebuah *non-frequent itemset* pasti juga akan merupakan *non-frequent*. Algoritma ini menerapkan pendekatan pencarian *frequent itemset* berdasarkan level, dimana *frequent k-itemset* digunakan untuk mencari *frequent (k+1)-itemset*. Awalnya dilakukan scan basis data untuk mencari *frequent 1-itemset* berdasarkan batasan nilai *min_supp*. Hasilnya berupa L_1 , kemudian L_1 digunakan untuk mencari L_2 , *frequent 2-itemset*, setelah itu L_2

digunakan untuk mencari L_3 , dan seterusnya sampai tidak ada lagi L_k atau *frequent k-itemset* yang bisa ditemukan. *Pseudocode* dari algoritma apriori ditunjukkan oleh Gambar 2.

```

L1 = {large 1-itemsets};
for { k = 2; Lk-1 ≠ ∅; k++ } do begin
  Ck = apriori-gen(Lk-1); // New candidates
  forall transactions t ∈ D do begin
    Ct = subset(Ck, t); Candidates contained in t
    forall candidates c ∈ Ct do
      c.count++;
  end
  Lk = {c ∈ Ck | c.count ≥ minsup}
end
Answer = ∪k Lk;

```

Gambar 2. Algoritma Apriori

Algoritma apriori bekerja dengan 2 tahap utama. Pada tahap yang pertama $L_{(k-1)}$ digunakan untuk membentuk C_k (*candidate generation*) dengan menggunakan fungsi *apriori-gen()*. Pada tahap yang kedua tiap – tiap *itemset* pada C_k akan dihitung nilai *support*-nya, hanya *itemset* yang memenuhi nilai *min_supp* yang akan digunakan untuk membentuk L_k .

Fungsi *apriori-gen()* terdiri dari 2 buah tahap yaitu *join* dan *prune* yang ditunjukkan oleh Gambar 3.

```

(Join)
insert into Ck
select p.item1, p.item2, ..., p.itemk-1, q.itemk-1
from Lk-1 p, Lk-1 q
where p.item1 = q.item1, ..., p.itemk-2 = q.itemk-2,
p.itemk-1 < q.itemk-1
(Prune)
forall itemsets c ∈ Ck do
  forall (k - 1)-subsets s of c do
    if (s ∉ Lk-1) then
      delete c from Ck;

```

Gambar 3. Join dan Prune

Pada tahap *join*, dilakukan penggabungan diantara $L_{(k-1)}$ dengan $L_{(k-1)}$ itu sendiri untuk mendapatkan C_k . Pada tahap *prune* dilakukan pemangkasan untuk menghapus $c \in C_k$ yang subsetnya tidak terdapat pada $L_{(k-1)}$. Contoh proses fungsi *apriori-gen()* dijelaskan sebagai berikut :

Misalkan diperoleh $L_3 = \{\{1\ 2\ 3\}, \{1\ 2\ 4\}, \{1\ 3\ 4\}, \{1\ 3\ 5\}, \{2\ 3\ 4\}\}$

Tahap *join* :

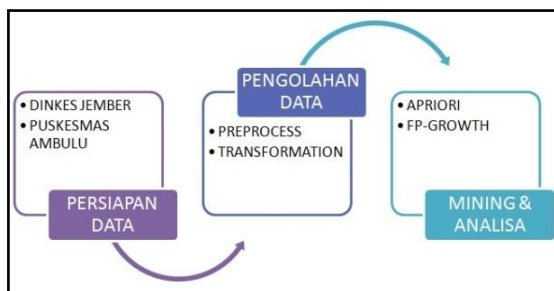
- 1) $\{1\ 2\ 3\}$ dikombinasikan dengan $\{1\ 2\ 4\}$ untuk membentuk $\{1\ 2\ 3\ 4\}$
- 2) $\{1\ 3\ 4\}$ dikombinasikan dengan $\{1\ 3\ 5\}$ untuk membentuk $\{1\ 3\ 4\ 5\}$
- 3) Sehingga, $C_4 = \{\{1\ 2\ 3\ 4\}, \{1\ 3\ 4\ 5\}\}$

Tahap *prune* :

- 1) Tiap-tiap 3-item subset dari $\{1\ 2\ 3\ 4\}$ akan diperiksa apakah semuanya terdapat di dalam L_3 , semua 3-item subset yang mungkin adalah $\{1\ 2\ 3\}$, $\{1\ 2\ 4\}$, $\{1\ 3\ 4\}$, dan $\{2\ 3\ 4\}$, dan ternyata keempatnya merupakan anggota L_3 , sehingga $\{1\ 2\ 3\ 4\}$ tidak dihapus dari C_4 . menjadi anggota L_4 , (*frequent 4-itemset*).
- 2) Tiap-tiap 3-item subset dari $\{1\ 3\ 4\ 5\}$ akan diperiksa apakah semuanya terdapat di dalam L_3 , semua 3-item subset yang mungkin adalah $\{1\ 3\ 4\}$, $\{1\ 3\ 5\}$, $\{1\ 4\ 5\}$, dan $\{3\ 4\ 5\}$, dan ternyata $\{1\ 4\ 5\}$ dan $\{3\ 4\ 5\}$ bukan merupakan anggota L_3 , sehingga $\{1\ 3\ 4\ 5\}$ dihapus dari C_4 .
- 3) Jadi didapat $C_4 = \{1\ 2\ 3\ 4\}$, sehingga L_4 (*frequent 4-itemset*) = $\{1\ 2\ 3\ 4\}$.

3. METODE PENELITIAN

Untuk dapat menemukan pola asosiasi keluhan pasien maka harus didapatkan terlebih dahulu data yang akan diproses, setelah pengumpulan data telah selesai maka akan dilakukan **preprocess** data kemudian dilanjutkan dengan **data transformation**, langkah selanjutnya yaitu melakukan **association rule mining** sehingga pola-pola asosiasi keluhan pasien bisa didapatkan. Rancangan alur penelitian ditunjukkan dalam Gambar 4, dan diikuti oleh penjelasan mengenai tahapan-tahapan penelitian.



Gambar 4. Rancangan Alur Penelitian

Pada penelitian ini dibagi menjadi 3 tahap utama yaitu tahap persiapan data, tahap pengolahan data, serta tahap mining dan analisa hasil keluaran.

Tahap Persiapan Data

Data yang digunakan pada penelitian ini adalah data nasabah pada suatu bank dan diambil dari laman <http://facweb.cs.depaul.edu/>. Data yang diambil tersebut termasuk data sekunder karena data ini merupakan hasil olahan pemilik *website*. Data yang diperoleh berbentuk file *.csv* (*comma separated value*), jumlah *record* sebanyak 600 buah, yang mempunyai 12 atribut. Tabel 2 dibawah ini merupakan contoh data yang diperoleh.

Data yang diperoleh ini tidak bisa langsung digunakan dalam proses *association rule mining*, sebelumnya data harus diperiksa apakah tidak ada data yang berulang, apakah terdapat duplikasi data, atau bahkan apakah ada data yang

tidak lengkap, selain itu diperiksa juga apakah dari sebelas atribut tersebut dapat dipakai atau tidak. Proses tersebut akan dilakukan pada tahap selanjutnya yaitu tahap pengolahan data, dimana nantinya akan ada 2 sub proses, *preprocess*, dan *data transformation*.

Tahap Pengolahan Data

Data awal yang sudah diperoleh kemudian akan dipersiapkan untuk proses *association rule mining* melalui tahap *preprocess* dan *transformation*. Kedua tahap ini menggunakan alat bantu yaitu *software* Microsoft Excel 2007 (*evaluation copy*). Penjelasan dua proses yang berurutan ini ditemukan dalam point-point berikut ini.

Tahap *Preprocess*, bertujuan untuk membersihkan dan memilih data. Membersihkan data dilakukan dengan memeriksa apakah ada duplikasi data, kesalahan pengetikan, serta data yang tidak lengkap. Sedangkan pemilihan data dilakukan dengan memilih variabel data yang diperlukan saja.

Tahap *Transformation*, bertujuan untuk merubah format data yang sudah ada menjadi format kategori sehingga bisa diproses lebih lanjut. Pada dataset yang dimiliki, atribut umur, dan pendapatan, yang berisi nilai numerik

Tabel 2. Contoh Data Nasabah

Id	age	sex	Region	income	married	children	car	save_act	current_act	mortgage	pep
ID12101	48	FEMALE	INNER_CITY	17546	NO	1	NO	NO	NO	NO	YES
ID12102	40	MALE	TOWN	30085.1	YES	3	YES	NO	YES	YES	NO
ID12103	51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES	YES	NO	NO
ID12104	23	FEMALE	TOWN	20375.4	YES	3	NO	NO	YES	NO	NO
ID12105	57	FEMALE	RURAL	50576.3	YES	0	NO	YES	NO	NO	NO
ID12106	57	FEMALE	TOWN	37869.6	YES	2	NO	YES	YES	NO	YES
ID12107	22	MALE	RURAL	8877.07	NO	0	NO	NO	YES	NO	YES
ID12108	58	MALE	TOWN	24946.6	YES	0	YES	YES	YES	NO	NO
ID12109	37	FEMALE	SUBURBAN	25304.3	YES	2	YES	NO	NO	NO	NO

akan dirubah berdasarkan kategori menjadi data nominal. Atribut umur nantinya akan digantikan oleh nilai numerik 0_34, 35_51, dan 52_max, yang merepresentasikan, usia muda, usia paruh baya, dan usia tua. Atribut pendapatan akan diisi nilai nominal 0_24386, 24387_43758, dan 43759_max, yang mewakili pendapatan rendah, pendapatan sedang, dan pendapatan tinggi.

Tahap Mining dan Analisa

Pencarian pola asosiasi diantara data nasabah bank akan dilakukan dengan algoritma apriori, dengan menentukan batasan nilai *support* dan *confidence* yang diinputkan oleh *user*. Setelah pola asosiasi didapatkan maka tahap yang terakhir adalah melakukan analisa terhadap *rule-rule* yang terbentuk, sehingga diharapkan akan ditemukan informasi-informasi baru yang menarik dan berguna.

Berdasarkan dataset yang diberikan *rule* yang terbentuk biasa dinotasikan dengan $X \rightarrow Y$ dimana X dan Y disini adalah *itemset*, bobot asosiasi disini berupa nilai *support* yang menjelaskan berapa kali sebuah *itemset* tercatat atau muncul dari sejumlah dataset dan nilai *confidence* yang menjelaskan seberapa kuat hubungan diantara *itemset* X dan Y

4. HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai hasil dari pencarian pola asosiasi data nasabah bank menggunakan algoritma apriori yang

meliputi tahap pengolahan data, serta tahap mining dan analisa. Masing-masing tahap tersebut akan dibahas dalam tiap-tiap poin yang sesuai dan disajikan pada uraian dibawah ini.

Tahap Pengolahan Data

Tahap pengolahan data adalah proses mempersiapkan dataset agar siap untuk proses pencarian pola asosiasi, data harus diperiksa apakah tidak ada data yang berulang, apakah terdapat duplikasi data, atau bahkan apakah ada data yang tidak lengkap. Selain itu diperiksa juga apakah dari kesebelas atribut tersebut dapat dipakai atau tidak. Dari Tabel 2 dapat kita lihat bahwa dataset terdiri dari 12 atribut, yaitu "Id", "age", "sex", "region", "income", "married", "children", "car", "save_act", "current_act", "mortgage", "pep". Masing-masing atribut tersebut mewakili keterangan nomor ID, faktor umur, jenis kelamin, domisili, pendapatan, status pernikahan, jumlah anak, kepemilikan mobil, kepemilikan rekening tabungan, kepemilikan rekening giro, kepemilikan hipotek, dan status pembelian *Personal Equity Plan* secara berurutan.

Dalam proses pencarian pola asosiasi maka data "Id" tidak diperlukan dalam proses *association rule mining* karena pola asosiasi yang dicari tidak berhubungan dengan informasi tentang nomor urut identitas nasabah, sehingga semua data pada kolom "Id" dihapus dan potongan tabel data dapat dilihat seperti pada Tabel 3.

Tabel 3. Dataset Hasil Pengolahan

Age	sex	region	income	Married	Children	car	save_act	current_act	mortgage	pep
48	FEMALE	INNER_CITY	17546	NO	1	NO	NO	NO	NO	YES
40	MALE	TOWN	30085.1	YES	3	YES	NO	YES	YES	NO
51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES	YES	NO	NO

23	FEMALE	TOWN	20375.4	YES	3	NO	NO	YES	NO	NO
57	FEMALE	RURAL	50576.3	YES	0	NO	YES	NO	NO	NO
57	FEMALE	TOWN	37869.6	YES	2	NO	YES	YES	NO	YES
22	MALE	RURAL	8877.07	NO	0	NO	NO	YES	NO	YES
58	MALE	TOWN	24946.6	YES	0	YES	YES	YES	NO	NO
37	FEMALE	SUBURBAN	25304.3	YES	2	YES	NO	NO	NO	NO

Tabel 4. Dataset Hasil Preproses

age	sex	Region	Income	Married	children	car	save_act	current_act	mortgage	pep
35_51	FEMALE	INNER_CITY	0_24386	NO	1	NO	NO	NO	NO	YES
35_51	MALE	TOWN	24387_43758	YES	3	YES	NO	YES	YES	NO
52_max	FEMALE	INNER_CITY	0_24386	YES	0	YES	YES	YES	NO	NO
0_34	FEMALE	TOWN	0_24386	YES	3	NO	NO	YES	NO	NO
52_max	FEMALE	RURAL	43759_max	YES	0	NO	YES	NO	NO	NO
52_max	FEMALE	TOWN	24387_43758	YES	2	NO	YES	YES	NO	YES
0_34	MALE	RURAL	0_24386	NO	0	NO	NO	YES	NO	YES
52_max	MALE	TOWN	24387_43758	YES	0	YES	YES	YES	NO	NO
35_51	FEMALE	SUBURBAN	24387_43758	YES	2	YES	NO	NO	NO	NO
52_max	MALE	TOWN	0_24386	YES	2	YES	YES	YES	NO	NO
52_max	FEMALE	TOWN	43759_max	YES	0	NO	YES	YES	NO	NO
52_max	FEMALE	INNER_CITY	24387_43758	NO	0	YES	YES	YES	YES	NO
35_51	FEMALE	TOWN	0_24386	YES	1	NO	YES	YES	YES	YES
52_max	FEMALE	TOWN	43759_max	YES	1	YES	YES	YES	YES	YES
35_51	MALE	RURAL	0_24386	YES	0	NO	YES	YES	YES	NO
35_51	FEMALE	INNER_CITY	0_24386	YES	0	YES	YES	YES	YES	NO
35_51	FEMALE	TOWN	0_24386	YES	2	NO	NO	NO	YES	NO
35_51	FEMALE	SUBURBAN	24387_43758	YES	0	NO	YES	NO	YES	NO
52_max	FEMALE	INNER_CITY	24387_43758	YES	0	NO	YES	NO	NO	YES

Tahap selanjutnya adalah melakukan transformasi data, yaitu merubah jenis data numerik ke dalam jenis nominal. Pada tabel diatas untuk data pada kolom "age" dan kolom "income" masih berjenis numerik sehingga akan dirubah ke dalam bentuk nominal. Untuk atribut "age" atau umur nantinya akan digantikan oleh nilai nominal 0_34, 35_51, dan 52_max, yang merepresentasikan, usia muda, usia paruh baya, dan usia tua. Atribut pendapatan akan diisi nilai nominal 0_24386, 24387_43758, dan 43759_max, yang mewakili pendapatan rendah, pendapatan sedang, dan pendapatan

tinggi. Tabel 4 menunjukkan sebagian hasil dari transformasi data yang telah dilakukan.

Tahap pengolahan data menghasilkan dataset yang siap untuk proses *association rule mining*. Secara kuantitas dataset tetap berisi 600 record dengan beberapa perubahan yang diperlukan. Pada sub bab berikutnya akan dijelaskan tentang proses *association rule mining* beserta analisa hasil pengujian.

Tahap Mining dan Analisa

Tahap proses *association rule mining* diimplementasikan menggunakan

perangkat lunak bantu WEKA, dimana di dalam WEKA sendiri sudah menyediakan bermacam-macam algoritma, dan algoritma apriori juga termasuk di dalamnya. Proses *association rule mining* ini akan mencari penemuan pola asosiasi dari kesebelas atribut yang diberikan, lebih khusus lagi akan dicari manakah atribut-atribut yang mempengaruhi seorang nasabah untuk membeli *Personal Equity Plan*.

Proses *association rule mining* pada penelitian ini akan menggunakan nilai *support* 5%, 10%, dengan nilai *confidence* yang digunakan adalah 90%. Dari keseluruhan *rule* yang terbentuk, maka akan diambil atau dipilih *rule* mana saja yang memuat *consequent* berupa atribut *Personal Equity Plan*.

Uji coba 1 (support 5% confidence 90%)

Total *rule* yang ditemukan adalah 371 buah, dari 371 buah *rule*, yang memenuhi *rule* yang dicari adalah sejumlah 60 buah. Ke-60 *rule* yang memenuhi tersebut disajikan pada Tabel 5.

Dari hasil uji coba 1 terdapat beberapa *rule* yang menarik. Pertama adalah 3 buah *rule* teratas yaitu :

1) **age=35_51 children=1 current_act=YES 35 => pep=YES 35 100%**, menunjukkan bahwa seorang

nasabah yang berumur antara 35 sampai 51 tahun dan mempunyai 1 orang anak serta mempunyai rekening giro akan memiliki kemungkinan 100% untuk ikut dalam program *Personal Equity Plan*.

2) **income=24387_43758 married=YES children=0 current_act=YES mortgage=NO 33 => pep=NO 33 100%**, menunjukkan bahwa seorang nasabah yang berpenghasilan antara US\$ 24.387 sampai US\$ 43.758, dengan status menikah, dan belum mempunyai anak, serta memiliki

rekening giro, dan tidak memiliki hipotek, mempunyai kemungkinan 100% untuk tidak ikut dalam program *Personal Equity Plan*.

3) **age=52_max married=YES children=0 save_act=YES current_act=YES 30 => pep=NO 30 100%**, menunjukkan bahwa seorang

nasabah yang berumur mulai 52 tahun keatas, dengan status menikah, dan belum memiliki anak, serta mempunyai rekening tabungan juga rekening giro, akan memiliki kemungkinan 100% untuk tidak ikut dalam program *Personal Equity Plan*.

Kedua adalah *rule* **married=YES children=0 current_act=YES mortgage=NO 88 => pep=NO 80 91%**,

rule ini menjelaskan bahwa kemungkinan 91% seorang nasabah tidak akan mengikuti program *Personal Equity Plan* apabila nasabah tersebut statusnya menikah, belum mempunyai anak, mempunyai rekening tabungan, dan tidak memiliki hipotek, pada *rule* ini juga dapat diketahui bahwa sebanyak 88% dari keseluruhan nasabah yang terdapat pada dataset mempunyai pola tersebut.

Uji coba 2 (support 10% confidence 90%)

Total *rule* yang ditemukan adalah 41 buah. Dari 41 buah *rule*, yang memenuhi *rule* yang dicari adalah sejumlah 4 buah. Ke-4 *rule* yang memenuhi tersebut disajikan pada Tabel 6.

Dengan batasan *support* 10% maka jumlah *rule* yang memenuhi lebih sedikit, seperti yang telah disajikan pada Tabel 6 maka hanya ada 4 buah *rule* yang memenuhi batasan *support* ini. Akan tetapi dari ke-4 *rule* tersebut tetap dapat diambil kesimpulan menarik tentang pola yang terbentuk. Pola asosiasi yang terbentuk dengan batasan *support* 10%

menghasilkan *rule-rule* dengan kondisi *antecedent* yang mempunyai nilai *support* diatas 50%. Bahkan dari ke-4 *rule* tersebut, paling kecil memiliki nilai *support* 70%, dan seperti yang telah diketahui pada umumnya, di dalam data mining lebih disukai *rule-rule* yang mempunyai nilai *support* yang tinggi.

Dari hasil uji coba dengan nilai *support* 10% maka diketahui pola asosiasi yang sering muncul adalah bahwa seorang nasabah memiliki kecenderungan tidak mengikuti program *Personal Equity Plan* apabila yang bersangkutan memenuhi salah satu dari keempat pola seperti dibawah ini :

- 1) Jika seorang nasabah statusnya menikah, tidak mempunyai anak, mempunyai rekening tabungan dan rekening giro maka kemungkinan 92% tidak akan mengikuti *Personal Equity Plan*.
- 2) Jika seorang nasabah menikah, tidak mempunyai anak, mempunyai rekening tabungan, dan tidak memiliki hipotek, maka kemungkinan 91% tidak akan mengikuti program *Personal Equity Plan*.
- 3) Jika seorang nasabah menikah, tidak mempunyai anak, memiliki rekening giro, dan tidak memiliki hipotek, maka

kemungkinan 91% tidak akan mengikuti program *Personal Equity Plan*.

- 4) Jika seorang nasabah berjenis kelamin wanita, menikah, tidak mempunyai anak, dan tidak memiliki hipotek, maka kemungkinan 90% tidak akan mengikuti program *Personal Equity Plan*.

Berdasarkan uji coba yang telah dilakukan, maka dapat kita ketahui bahwa seorang nasabah akan memiliki kemungkinan 100% untuk ikut dalam program *Personal Equity Plan* jika berumur antara 35 sampai 51 tahun dan mempunyai 1 orang anak serta mempunyai rekening giro. Sebaliknya, seorang nasabah kemungkinan 100% untuk tidak ikut dalam program *Personal Equity Plan* jika berpenghasilan antara US\$ 24.387 sampai US\$ 43.758, dengan status menikah, dan belum mempunyai anak, serta memiliki rekening giro, dan tidak memiliki hipotek. Selain itu seorang nasabah juga memiliki kemungkinan 100% tidak ikut dalam program *Personal Equity Plan* jika berumur mulai 52 tahun keatas, dengan status menikah, dan belum memiliki anak, serta mempunyai rekening tabungan juga rekening giro.

Tabel 5. Hasil *Association rule mining* (*support* 5%)

NO	RULE	CONFIDENCE
1	age=35_51 children=1 current_act=YES 35 ==> pep=YES 35	100%
2	income=24387_43758 married=YES children=0 current_act=YES mortgage=NO 33 ==> pep=NO 33	100%
3	age=52_max married=YES children=0 save_act=YES current_act=YES 30 ==> pep=NO 30	100%
4	income=24387_43758 married=YES children=0 mortgage=NO 44 ==> pep=NO 43	98%
5	income=24387_43758 married=YES children=0 save_act=YES current_act=YES 33 ==> pep=NO 32	97%
6	married=NO children=0 save_act=YES mortgage=NO 32 ==> pep=YES 31	97%
7	age=35_51 married=YES children=1 31 ==> pep=YES 30	97%
8	income=24387_43758 children=1 mortgage=NO 31 ==> pep=YES 30	97%
9	income=24387_43758 married=YES children=0 save_act=YES mortgage=NO 31 ==> pep=NO 30	97%
10	age=35_51 children=1 48 ==> pep=YES 46	96%
11	age=35_51 married=YES children=0 save_act=YES 44 ==> pep=NO 42	95%

12	age=35_51 children=1 save_act=YES 38 ==> pep=YES 36	95%
13	married=NO children=0 current_act=YES mortgage=NO 37 ==> pep=YES 35	95%
14	sex=FEMALE married=YES children=0 save_act=YES current_act=YES mortgage=NO 37 ==> pep=NO 35	95%
15	age=35_51 married=YES children=0 current_act=YES mortgage=NO 35 ==> pep=NO 33	94%
16	married=NO children=0 mortgage=NO 48 ==> pep=YES 45	94%
17	income=24387_43758 children=1 save_act=YES current_act=YES 32 ==> pep=YES 30	94%
18	age=35_51 married=YES children=0 save_act=YES current_act=YES 32 ==> pep=NO 30	94%
19	age=35_51 married=YES children=0 mortgage=NO 47 ==> pep=NO 44	94%
20	income=24387_43758 married=YES children=0 save_act=YES 47 ==> pep=NO 44	94%
21	sex=FEMALE married=YES children=0 save_act=YES current_act=YES 47 ==> pep=NO 44	94%
22	children=0 save_act=YES current_act=YES mortgage=YES 46 ==> pep=NO 43	93%
23	married=YES children=0 save_act=YES current_act=YES mortgage=NO 61 ==> pep=NO 57	93%
24	age=52_max children=1 43 ==> pep=YES 40	93%
25	income=24387_43758 children=1 save_act=YES 41 ==> pep=YES 38	93%
26	income=24387_43758 children=1 current_act=YES 41 ==> pep=YES 38	93%
27	region=TOWN married=YES children=0 save_act=YES 40 ==> pep=NO 37	93%
28	income=24387_43758 children=1 53 ==> pep=YES 49	92%
29	married=YES children=0 car=YES save_act=YES current_act=YES 39 ==> pep=NO 36	92%
30	married=YES children=0 save_act=YES current_act=YES 87 ==> pep=NO 80	92%
31	children=0 save_act=YES mortgage=YES 62 ==> pep=NO 57	92%
32	children=0 save_act=NO mortgage=YES 37 ==> pep=YES 34	92%
33	married=YES children=0 car=YES save_act=YES mortgage=NO 37 ==> pep=NO 34	92%
34	married=YES children=0 car=YES current_act=YES mortgage=NO 37 ==> pep=NO 34	92%
35	married=YES children=1 save_act=YES current_act=YES 48 ==> pep=YES 44	92%
36	sex=FEMALE married=YES children=0 save_act=YES mortgage=NO 48 ==> pep=NO 44	92%
37	married=YES children=0 car=NO save_act=YES current_act=YES 48 ==> pep=NO 44	92%
38	age=52_max children=1 save_act=YES 36 ==> pep=YES 33	92%
39	sex=MALE children=0 save_act=YES mortgage=YES 36 ==> pep=NO 33	92%
40	sex=FEMALE married=YES children=0 car=NO save_act=YES 36 ==> pep=NO 33	92%
41	region=INNER_CITY married=YES children=0 current_act=YES mortgage=NO 46 ==> pep=NO 42	91%
42	married=YES children=0 save_act=YES mortgage=NO 80 ==> pep=NO 73	91%
43	region=TOWN married=YES children=0 current_act=YES 34 ==> pep=NO 31	91%
44	income=0_24386 children=2 save_act=YES current_act=YES 34 ==> pep=NO 31	91%
45	sex=FEMALE region=INNER_CITY married=YES children=0 mortgage=NO 34 ==> pep=NO 31	91%
46	married=YES children=0 car=NO save_act=YES current_act=YES mortgage=NO 34 ==> pep=NO 31	91%
47	married=YES children=0 current_act=YES mortgage=NO 88 ==> pep=NO 80	91%
48	sex=FEMALE married=YES children=0 current_act=YES mortgage=NO 55 ==> pep=NO 50	91%
49	income=0_24386 children=2 car=YES 33 ==> pep=NO 30	91%
50	age=52_max married=YES children=0 mortgage=NO 33 ==> pep=NO 30	91%
51	sex=MALE married=YES children=0 current_act=YES mortgage=NO 33 ==> pep=NO 30	91%
52	married=YES children=0 car=NO save_act=YES mortgage=NO 43 ==> pep=NO 39	91%

53	married=YES children=0 car=NO save_act=YES 64 ==> pep=NO 58	91%
54	sex=FEMALE married=YES children=0 save_act=YES 63 ==> pep=NO 57	90%
55	age=52_max married=YES children=0 save_act=YES 41 ==> pep=NO 37	90%
56	sex=FEMALE married=YES children=0 car=NO mortgage=NO 41 ==> pep=NO 37	90%
57	married=YES children=0 car=NO current_act=YES mortgage=NO 51 ==> pep=NO 46	90%
58	sex=FEMALE married=YES children=0 mortgage=NO 70 ==> pep=NO 63	90%
59	age=35_51 married=YES children=0 car=NO 40 ==> pep=NO 36	90%
60	sex=MALE married=YES children=0 save_act=YES current_act=YES 40 ==> pep=NO 36	90%

Tabel 6. Hasil *Association rule mining* (support 10%)

NO	RULE	CONFIDENCE
1	married=YES children=0 save_act=YES current_act=YES 87 ==> pep=NO 80	92%
2	married=YES children=0 save_act=YES mortgage=NO 80 ==> pep=NO 73	91%
3	married=YES children=0 current_act=YES mortgage=NO 88 ==> pep=NO 80	91%
4	sex=FEMALE married=YES children=0 mortgage=NO 70 ==> pep=NO 63	90%

5. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini sebagai berikut.

- 1) Seorang nasabah akan memiliki kemungkinan 100% untuk ikut dalam program *Personal Equity Plan* jika berumur antara 35 sampai 51 tahun dan mempunyai 1 orang anak serta mempunyai rekening giro.
- 2) Seorang nasabah kemungkinan 100% untuk tidak ikut dalam program *Personal Equity Plan* jika berpenghasilan antara US\$ 24.387 sampai US\$ 43.758, dengan status menikah, dan belum mempunyai anak, serta memiliki rekening giro, dan tidak memiliki hipotek.
- 3) Seorang nasabah juga memiliki kemungkinan 100% tidak ikut dalam program *Personal Equity Plan* jika berumur mulai 52 tahun keatas, dengan status menikah, dan belum memiliki anak, serta mempunyai rekening tabungan juga rekening giro.

Adapun saran pengembangan penelitian ini adalah :

- 1) Belum adanya fitur antar muka yang memudahkan user dalam pengoperasian data mining.
- 2) Pengembangan dapat dilakukan dengan menambahkan fitur pendukung keputusan seperti penentuan nasabah mana yang perlu diutamakan dalam pelayanan, penentuan nasabah yang tidak mempunyai prospek, dan visualisasi berbasis peta geografis untuk menunjukkan segmen nasabah.

DAFTAR PUSTAKA

- Witten, I. H., Frank, E., *Data Mining Practical Machine Learning Tools and Technique*, Morgan Kaufmann Publishers, 2005.
- Sumathi, S., Sivanandam, S. N., *Introduction to Data Mining and its Applications*, Springer, 2006.
- Danapana, H., Roy, M. S., *Effective Data Mining Association Rules for Heart Disease Prediction System*, IJCST Vol. 2, October – December, 2011.
- Srinivas, K., Rao, G. R., Govardhan, A., *Mining Association Rules from Large Datasets Towards Disease Prediction*, International Conference on Information and Computer

- Networks, 2012.
- Ordonez, C., Santana, C. A., de Braal, L., Discovering Interesting Association Rules in Medical Data, Proceedings of ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 2000, Hal. 78 – 85.
- Erwin, Analisis Market Basket Dengan Algoritma Apriori dan FP-Growth, Jurnal Generic Vol. 4 No. 2, Juli 2009.
- Han, J., Kamber, M., Pei, J., Data Mining Concepts and Techniques Third Edition, Morgan Kaufmann Publisher, 2012
- Han, J., et al, Mining Frequent Pattern Without Candidate Generation A Frequent-Pattern Tree Approach, Data Mining and Knowledge Discovery, 8, 53–87, 2004.