BAB IV
PEMAHAMAN ISU BISNIS

Langkah awal dalam sebelum memulai *data mining* adalah bagaimana mengidentifikasikan permasalahan bisnis yang dihadapi dan bagaimana menterjemahkan permasalahan bisnis tersebut kedalam sekumpulan pertanyaan yang bisa diselesaikan oleh *data mining*.

4.1 Gambaran umum industri perbankan indonesia

Bank sebagai tempat menyimpan dana yang aman dan menguntungkan dengan tingkat risiko rendah pada beberapa tahun belakangan ini mengalami gejolak yang luar biasa. Sejak krisis ekonomi melanda Indonesia, Bisnis perbankan konsumer sebagai layanan perbankan kepada perseorangan berkembang pesat, dipicu oleh banyaknya pelaku ekonomi yang mengalami masalah serius dengan kondisi keuangannya (*financial distress*). Terpuruknya bisnis *corporate banking* membuat pangsa pasar bisnis ini menyempit, lalu mendorong beberapa bank mengalihkan target pasar ke pangsa individual/perorangan yang dikenal dengan bisnis perbankan konsumer.

Persaingan yang semakin ketat, baik sesama bank dalam hal produk, promosi dan suku bunga. Juga persaingan dengan perusahaan pembiayaan dan alternatif pembiayaan lain seperti pegadaian, koperasi, dan Lembaga Perkreditan Desa (LPD) dan besarnya tuntutan nasabah sejalan dengan peningkatan 'banking *mindedness*' yang menuntut perbankan terus melakukan pengembangan produk, kualitas layanan dan teknologi informasi (TI).

Perkembangan bisnis perbankan konsumer mengarahkan bank – bank nasional untuk meningkatkan jumlah *customer base* atau nasabah, dengan cara mempertahankan nasabah lama dan menambah nasabah baru melalui berbagai hal misalnya pengembangan jaringan, inovasi produk, program dan berbagai hal lainnya.

Jumlah nasabah atau *customer base* merupakan hal dasar dalam mengembangkan dan menghadapi persaingan bisnis perbankan. Ibu Dewi, seorang *Departement Head* sebuah bank terkemuka di Indonesia menjawab beberapa alasan mengapa jumlah nasabah begitu penting bagi perkembangan bisnis perbankan? [13]

Pertama, perusahaan akan memiliki kapasitas untuk melakukan cross-selling. Dengan semakin tingginya *wallet share* karena keberhasilan melakukan *cross-selling*, justru pelanggan akan lebih puas dan lebih loyal. Mereka cenderung melakukan interaksi yang lebih banyak dan pelanggan juga lebih tergantung pada perusahaan. Seorang nasabah yang memiliki tabungan dan deposito, kartu kredit, KPR serta giro, biasanya akan lebih puas terhadap bank dimana dia menjadi nasabahnya.

Kedua, perusahaan akan mampu memberikan nilai tambah yang lebih besar. Ini bisa terjadi karena perusahaan memiliki bargaining power yang tinggi ketika bernegosiasi dengan pihak-pihak lain seperti supplier atau rekanan. Mereka cenderung mendapatkan harga yang kompetitif dari pihak supplier dan akibatnya, mampu memberikan harga yang kompetitif pula kepada pelanggannya. Perusahaan yang memiliki jumlah pelanggan besar, juga mampu mendapatkan dukungan lebih besar dari rekanan atau pihak-pihak yang beraliansi dengan perusahaan tersebut.

Pada akhirnya, perusahaan akan mampu memberikan nilai tambah dalam bentuk kualitas produk dan layanan yang lebih tinggi pula.

Ibu Dewi menambahkan juga, bahwa dalam konteks *customer base*, yang perlu diperhatikan adalah bukan hanya jumlah tetapi juga kualitas. Pelanggan yang sudah lama menggunakan produk atau layanan dari perusahaan tersebut dan pelanggan yang memilki daya beli yang tinggi, adalah pelanggan yang cenderung memiliki kualitas lebih baik. Perusahaan yang mampu untuk mengangkat kekuatan *customer base* inilah yang akhirnya, mampu memberikan nilai tambah kepada pelanggannya untuk kepuasan pelanggan.

4.2 Persaingan ketat dalam industri perbankan

Persepsi masyarakat pengguna jasa Bank terhadap Bank tidak akan lepas dari image yang melekat pada Bank tersebut yang dikaitkan dengan pelayanannya kepada nasabah, berbagai jenis produk/jasa, dan kemampuan teknologi informasi pendukungnya, yang pada akhirnya paduan dari ketiga hal tersebut akan menjadi daya tarik kepada nasabah. Persaingan dalam memenuhi tuntutan nasabah yang pada intinya adalah dipenuhinya 'kemudahan melakukan transaksi perbankan dimana saja dan kapan saja', turut mendorong persaingan teknologi perbankan, yang dapat dicontohkan dengan adanya kemampuan feature *online realtime* di seluruh cabangnya, berbagai *delivery channel* bermuatan teknologi untuk berhubungan dengan para nasabah yaitu: ATM (Automatic Teller Machine), Telephone Banking, PC Banking, *Internet Banking*, *TV Banking*, dan *Mobile Banking*. Kemajuan TI di industri finansial memungkinkan Bank untuk melakukan '*leap-frog*' dalam memenuhi kecenderungan tuntutan nasabah akan kemudahan

akses, kenyamanan dan ketersediaan-nya setiap saat. Bank-bank yang dapat memanfaatkan teknologi informasi secara cerdas untuk mendukung produk dan layanannya akan memenangkan persaingan tersebut.

Guna meningkatkan jumlah nasabah, saat ini bank – bank berlomba untuk meningkatkan kualitas layanan seperti pengembangan jaringan dan melakukan inovasi produk dan program khususnya untuk nasabah konsumer. Namun trend yang berkembang dikalangan perbankan saat ini adalah ketika salah satu bank meluncurkan inovasi produk atau program promosi baru maka bank pesaing akan mengeluarkan inovasi produk atau program promosi yang secara garis besar sama namun memberikan keuntungan yang lebih menarik kepada nasabah, salah satu contoh konkrit adalah ketika Bank Mandiri meluncurkan program promosi undian berhadiah mobil maka dalam kurun waktu yang kurang lebih sama Bank BNI dan beberapa bank lainnya juga meluncurkan program promosi yang sama. Contoh konkrit lainnya adalah ketika Bank BII meluncurkan produk Tabungan Gold maka beberapa bank juga meluncurkan produk serupa antara lain Tahapan Gold BCA, Tabungan Gold Mandiri dan berbagai produk serupa lainnya.

Inovasi atau program promosi ini diluncurkan guna memberikan apresiasi kepada nasabah lama sehingga dapat dipertahankan dan untuk menarik nasabah baru dari bank lain. Namun peluncuran berbagai inovasi program, produk dan pengembangan jaringan tidak sembarangan dilakukan oleh para pelaku perbankan, sebelum meluncurkan perihal tersebut, terdapat aktifitas pemahaman karakteristik nasabah dan kebutuhan nasabah yang atau dengan kata lain para pelaku perbankan terlebih dahulu melakukan prediksi sebagai berikut karakteristik nasabah X akan melakukan tindakan Y apabila dihadapkan pada suasana Z.

BAB V
PEMAHAMAN DAN PERSIAPAN DATA

Data mining tidak akan pernah dapat dilakukan tanpa tersedianya data yang akan di tambang. Untuk itu dibutuhkan pemahanan akan data yang seperti apa yang akan dipakai dalam proses *data mining* nanti. Pada bab ini akan dibahas mengenai data seperti apa yang tersedia dan bagaimana menggunakan data yang tersedia tersebut agar bisa dianalisa.

Data yang akan dipakai dalam proses *data mining*, perlu dipersiapkan dan diubah kedalam bentuk model yang lebih sederhana agar bisa dianalisa dengan memakai teknik *data mining*. Data yang dipakai secara umum bisa dibagi kedalam empat jenis data, yaitu:

- Data demografis nasabah
- Data relationship
- Data transaksional
- Data tambahan

Data demografis adalah data yang berkaitan dengan nasabah secara individual, seperti tempat tinggal, umur, jenis kelamin dan lain sebagainya. Data demografis disa disajikan lebih baik dengan bantuan tambahan dari sumber data yang lain. Seperti data mengenai kabupaten, kecamatan, maupun provinsi.

Data relationship menjabarkan mengenai bagaimana hubungan antara pihak bank dan nasabahnya, seperti data mengenai melalui media apa seorang nasabah

terdaftar, kapan pertama kali nasabah tersebut berhubungan dengan pihak bank, siapa yang merekomdasikan nasabah tersebut.

Data transaksional merupakan data mengenai jumlah dan ukuran transaksi yang dilakukan oleh nasabah secara individual. Data ini bisa berbentuk data historis mengenai fluktuasi naik turun jumlah saldo yang dimiliki oleh nasabah. Dikarenakan data yang berbentuk historis, maka data transaksional butuh diolah melalui teknik agregasi data.

Data tambahan merupakan data yang diambil dari sumber data lain. Data ini bisa berasal dari dalam maupun dari luar institusi. Data dari dalam institusi bisa berupa data segmentasi nasabah, data jenis produk. Data dari luar institusi biasanya tersedia secara umum dan bisa dipakai untuk berbagai kebutuhan seperti data mengenai struktur pembagian wilayah sebuah provinsi.

5.1 *Dataset* awal

Dataset awal adalah *dataset* yang diambil langsung dari *datamart* institusi bank yang di jadikan obyek penelitian ini. *Dataset* tersebut dipersiapkan oleh staff terkait dalam bentuk dua buah tabel sejenis yaitu:

- Tabel Close yaitu sekumpulan *dataset* historis dari nasabah-nasabah yang sudah menutup rekening mereka. Tabel ini memiliki jumlah *record* sebanyak 5025 *record.*
- Tabel Open yaitu sekumpulan *dataset* dari nasabah-nasabah yang masih dengan rekening aktif. Tabel ini berisikan record sebanyak 5000 *record.*

Kedua tabel tersebut memiliki atribut-atribut yang identik satu dengan yang lainnya. Atribut-atribut tersebut antara lain.

Nama atribut	Keterangan
REGNO	Atribut Referensi
BRANCH CODE	Kode Cabang pembuka Rekening
SEX	Jenis Kelamin Pemilik Rekening (Male, Female, N/A)
PRODUCT	Jenis Produk yang dipilih (Tabungan, Tabungan Berjangka, Giro)
CUST.TYPE	Segmentasi nasabah berdasarkan Tipe Nasabah (Individual, Corporate)
CUST. CLASS	Segmentasi nasabah berdasarkan kelas nasabah (Reguler, Gold, Platinum)
HasATMCard	Nasabah memakai fasilitas ATM (Yes, No)
REFERALBY	Jabatan staff bank yang mereferensikan nasabah ketika membuka rekening (nilai variabel #N/A# menandakan nasabah datang sendiri ketika membuka rekening)
JOIN DATE	Tanggal nasabah pertama kali bergabung
OPEN DATE	Tanggal Rekening dibuka
CLOSE DATE	Tanggal Rekening ditutup
ADRESS1	Alamat tempat tinggal nasabah
ADRESS2	Alamat tempat tinggal nasabah
ADRESS3	Alamat tempat tinggal nasabah
BALANCE DATE	Data historis saldo harian rata-rata dari rekening nasabah

Tabel 2. Atribut *dataset* awal dan keterangannya

Tabel *dataset* awal tersebut merupakan data mentah yang diambil langsung dari *datamart*. Bentuk tabel dataset awal yang horisontal mempersulit proses analisa lebih lanjut. Oleh karena itu dibutuhkan transformasi tabel ke dalam bentuk vertikal.

Kualitas dari data yang didapat secara umum sudah cukup bagus kecuali untuk atrribut-atribut tertentu yang sangat tidak konsisten dalam hal penyajian datanya. Inkonsistensi tersebut antara lain:

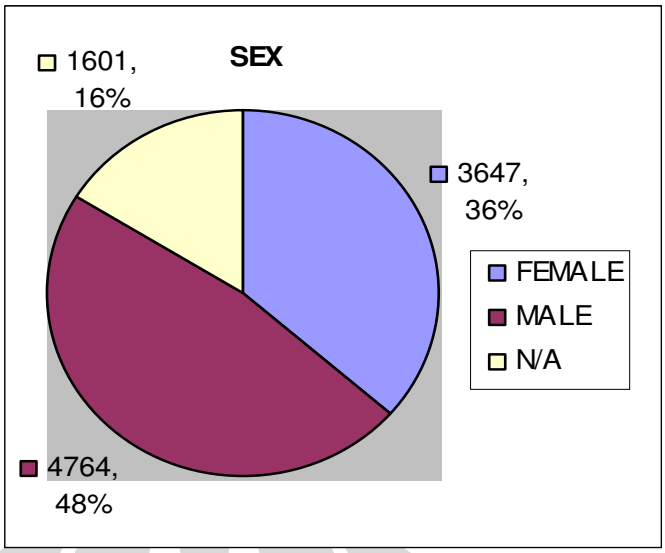
- *Missing value,* yaitu nilai yang yang hilang dalam sebuah variabel atau dengan kata lain variabel tersebut tidak bernilai.
- *Data noise,* yaitu data-data yang penulisannya tidak seragam.

5.2 Eksplorasi *dataset* awal

Ekplorasi data dibutuhkan untuk memberi gambaran awal seperti apa data yang akan dipakai. Dengan melakukan eksplorasi data awal, bisa didapatkan bentuk sebaran data dalam masing-masing *variabel* serta bisa dideteksi nilai-nilai yang tidak valid dalam kumpulan data tersebut seperti *missing value*, sehingga memungkinkan dilakukannya pembersihan data.

5.2.1 Variabel sex

Variabel Sex adalah variabel yang berisi informasi mengenai jenis kelamin pemilik rekening. Variabel ini terbagi atas tiga buah kategori yaitu Female, Male, dan N/A. Kategori N/A adalah kategori merupakan *missing value* yang dimiliki oleh variabel ini.



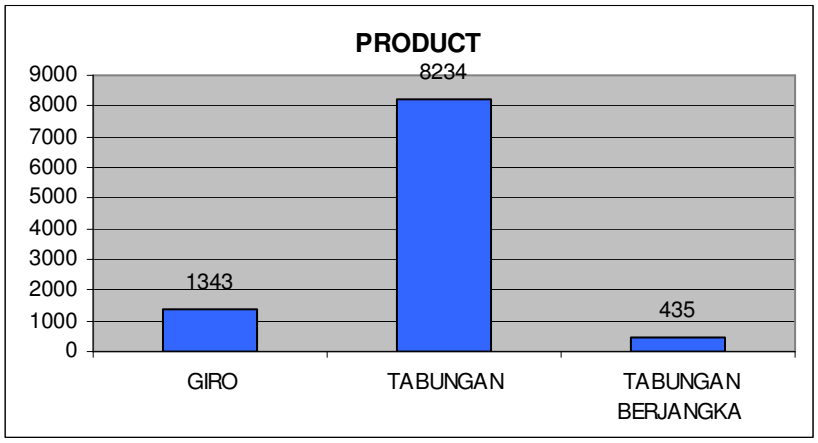
Gambar 9. Distribusi data variabel Sex

Distribusi data variabel Sex ini seperti yang terlihat pada gambar 9 sebagian besar berada pada memiliki jenis kelamin laki-laki (48%). Sedangkan 36% terdiri dari pemilik rekening dengan jenis kelamin perempuan (*female*) sedangkan sisanya sebesar 16% adalah data-data yang tidak memiliki nilai (*missing value*)

5.2.2 *Variabel* Product

Variabel product merupakan *variabel* yang mengkategorikan jenis product simpanan perbankan yang digunakan oleh nasabah. Walaupun produk simpanan perbankan yang ditawarkan terdiri dari banyak jenis produk, akan tetapi produk-produk tersebut bisa dikategorikan ke dalam tiga jenis produk, yaitu Giro, Tabungan, dan Tabungan berjangka.

Dari distribusi data *variabel* product seperti terlihat pada gambar 10, terlihat bahwa mayoritas sample nasabah memakai produk tabungan (82,24%), lalu diikuti oleh produk Giro (13,41%), dan sisanya memakai produk tabungan berjangka.

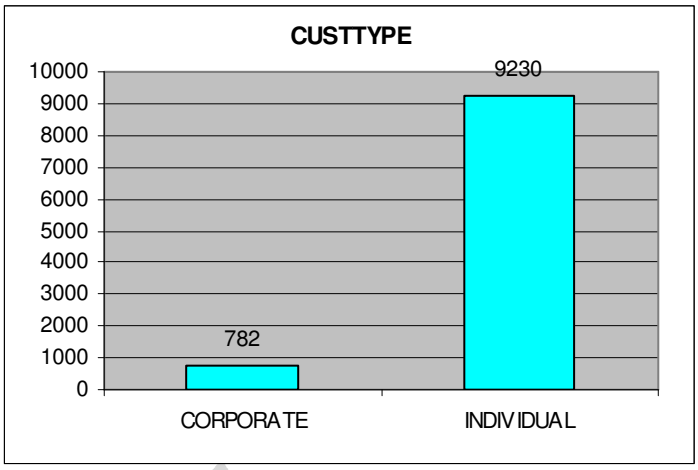


Gambar 10. Distribusi data variabel Product

5.2.3 Variabel Custtype

Variabel *Custtype* adalah variabel yang mengkategorikan jenis kepemilikan sebuah rekening. Kepemilikan rekening ini terbagi atas dua jenis yaitu rekening individual dan rekening korporat. Rekening individual adalah rekening yang dimiliki atau diatasnamakan oleh nasabah-nasabah perseorangan, sedangkan rekening korporat adalah rekening-rekening yang kepemilikiannya diatasnamakan kepada sebuah institusi, baik itu berupa perusahaan, lembaga, maupun badan usaha yang lain.

Dari 10012 sample rekening yang dijadikan penelitian, 9230 sample atau sekitar 92,1% adalah rekening dengan jenis kepemilikan individual, sedangkan sisanya sebanyak 7.9% adalah rekening dengan kepemilikan bertipe korporat. Grafik distribusi data variabel *custtype* dapat dilihat pada gambar 11 dibawah ini.

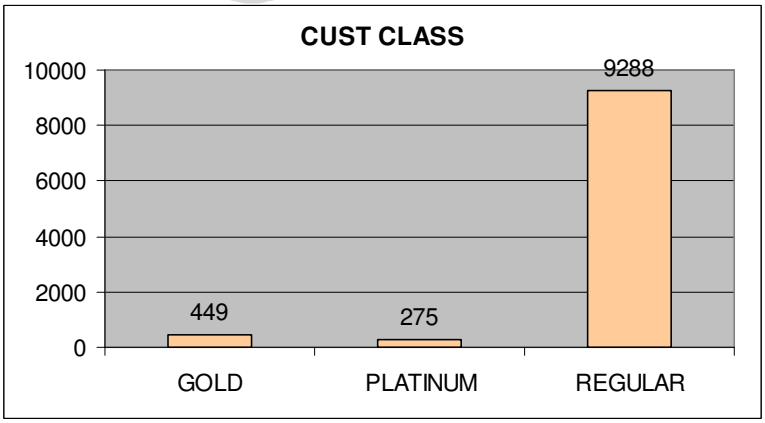


Gambar 11. Distribusi variabel custtype

5.2.4 Varibel CustClass

Variabel *Cust Class* merupakan variabel yang menunjukkan kelas atau jenis simpanan yang dimiliki oleh seorang nasabah. Seorang nasabah ditentukan kelasnya berdasarkan jumlah simpanan yang dimiliki oleh nasabah tersebut. Variabel *CustClass* ini mengelompokan nasabah kedalam tiga jenis kelas, yaitu kelas Gold, Platinum, dan Regular.

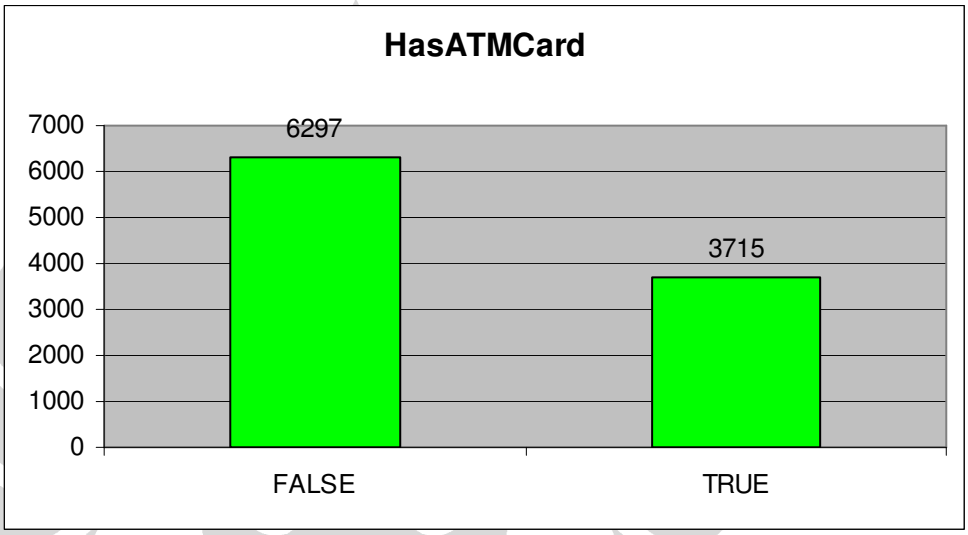
Distribusi data variabel *CustClass* seperti yang terlihat pada gambar 12 menunjukan bahwa mayoritas sample data yang akan dipakai berada pada kelas *Regular* (92,77%) sedangakan sisanya berada pada kelas Gold (4,48%) dan kelas platinum (2,74%).



Gambar 12. Distribusi Variabel CustClass

5.2.5 Varibel HasATMCard

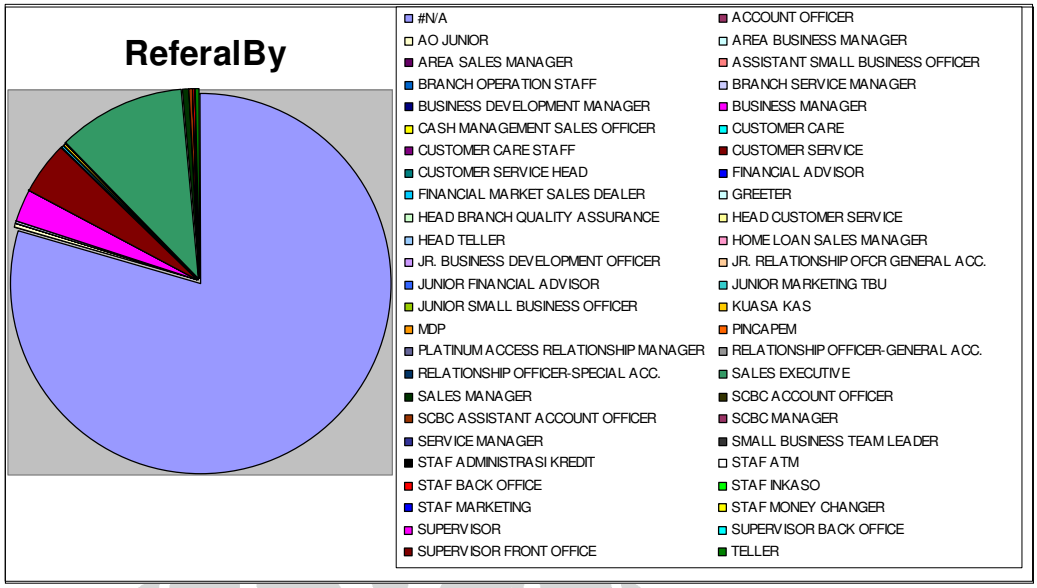
Variabel HasATMCard merupakan variabel yang menunjukkan status kepemilikan kartu ATM oleh seorang nasabah. *Variabel* ini hanya terdiri atas dua nilai yaitu *true* dan *false*. Seperti terlihat pada gambar 13, dari keseluruhan sample yang digunakan, sekitar 37,10% memiliki kartu ATM dan sisanya 62,90% tidak memiliki Kartu ATM.



Gambar 13. Rasio Kepemilikan Kartu ATM

5.2.6 Variabel ReferalBy

Variabel *ReferalBy* merupakan variabel yang memetakan melalui siapa seorang nasabah direkomendasikan ketika menjadi customer. Tersebar ke dalam 50 jenis tipe yang data yang berbeda. Sebagian besar data yang ada adalah data yang tidak memiliki nilai. Sedangkan sisanya merupakan kumpulan data dari berbagai jenis jabatan.



Gambar 14. Sebaran data variabel ReveralBy

5.3 Persiapan data

Bentuk tabel dataset awal yang masih berbentuk horizontal menjadikan tabel tersebut sulit untuk dimodifikasi dan ditangani lebih lanjut. Untuk mengatasi hal tersebut, tabel dataset awal dipecah menjadi 2 buah tabel, dengan nama tabel acc_info dan tabel balance.

Dalam tabel acc_info secara umum terbagi atas 2 jenis data, yaitu:

- Data demografis yang berupa data mengenai jenis kelamin nasabah dan data mengenai tempat tinggal nasabah.
- Data mengenai segmentasi nasabah, dan data mengenai tipe produk yang dipakai oleh nasabah.
- Data aggregasi atau data turunan, yaitu data yang diturunkan dari data awal. Data ini pada umumnya merupakan data hasil perhitungan dari beberapa *variabel* yang bisa dihitung. Sebagai contoh, varibel tenureweek pada tabel 3 dibawah merupakan *variabel* yang didapat

dengan mengurangi variabel CLOSE DATE dengan *variabel* OPEN DATE dalam satuan minggu.

Nama atribut	Keterangan
REGNO	Atribut Referensi
BRANCH CODE	Kode Cabang pembuka Rekening
SEX	Jenis Kelamin Pemilik Rekening (Male, Female, N/A)
PRODUCT	Jenis Produk yang dipilih (Tabungan, Tabungan Berjangka, Giro)
CUSTTYPE	Klasifikasi Tipe Nasabah (Individual, Corporate)
CUST CLASS	Klasifikasi kelas nasabah (Reguler, Gold, Platinum)
DATI II	Dati II tempat nasabah berada
PROVINCE	Propinsi tempat nasabah berada
HasATMCard	Nasabah memakai fasilitas ATM (Yes, No)
ReferalBy	Jabatan staff bank yang mereferensikan nasabah ketika membuka rekening (nilai variabel #N/A# menandakan nasabah datang sendiri ketika membuka rekening)
Voluntairly_join	Flag bilamana nasabah secara sukarela (voluntair) membuka rekening (yes, no). Diturunkan dari variabel ReferalBy, dimana nilai #N/A# dari variabel adalah nilai “yes” dari variabel Voluntairly_join.
JoinDate	Tanggal nasabah pertama kali menjadi nasabah
Dtopen	Tanggal rekening dibuka (1 – 31)
Mtopen	Bulan rekening dibuka (1 – 12)
Yropen	Tahun rekening dibuka
OPEN DATE	Tanggal Rekening dibuka
CLOSE DATE	Tanggal Rekening ditutup
Week_open	Minggu rekening dibuka (1-52)
Tenuredays	Umur rekening dalam satuan hari
Tenureweek	Umur rekening dalam satuan minggu
Tenuremonth	Umur rekening dalam satuan bulan
Close_status	Flag status rekening, apakah sudah ditutup atau masih berjalan (yes, no)
Rec_type	Digunakan untuk membedakan data yang dipakai untuk pembentukan model (”train”) dan data yang dipakai untuk menguji akurasi model (”test”).

Tabel 3. Tabel Acc_info

Pada tabel acc_info tersebut ditambahkan juga beberapa variabel tambahan, antara lain atribut *regno* yang dijadikan sebagai primary key dari tabel acc_info,

variabel *Province, Dati II, Voluntairly_join, dtopen, mtopen, yropen, week_open, tenureday, tenureweek, tenuremonth, Close_status* dan atribut *rec_type.*

5.3.1 Atribut *Province* dan *Dati II*

Atribut Province dan Dati II merupakan hasil generalisasi dari atribut alamat1, alamat2, dan alamat3 pada *dataset* awal. Kesulitan yang dihadapi pada pembentukan atribut province dan Dati II tersebut adalah data yang proses pemberian nilai dilakukan secara manual satu persatu, selain itu data yang terekam pada atribut *dataset* awal sangat beragam dan tidak konsisten. Bentuk ketidakragaman tersebut antara lain:

- Alamat pada *dataset* awal tidak mencantumkan kode pos, tapi hanya mencantumkan nama Dati II.
- Alamat tidak mencantumkan Dati II maupun Propinsi akan tetapi mencantumkan nama kelurahan dan kode pos.
- Alamat hanya mencantumkan nama jalan saja tanpa menyebutkan nama Dati II maupun KodePos.

Untuk alamat yang mencantumkan kode pos dapat dengan mudah diubah dengan membandingkan kode pos tersebut dengan data eksternal yang didapat dari data kodepos yang dimiliki oleh kantor pos nasional. Sedangkan alamat yang hanya mencantumkan nama Dati II dapat diubah dengan membandingkan dengan data yang dimiliki oleh BPS.

5.3.2 Atribut *dtopen, mtopen, yropen, week_open*

Atribut *dtopen, mtopen, yropen, week_open* merupakan atribut tambahan yang diturunkan dari atribut *open date* dari *dataset* awal. Atribut-atribut tersebut

tidak sulit dilakukan karena karakteristik dari atribut *open date* yang sudah memakai format dd/mm/yyyy. Sedangkan nilai untuk atribut *week_open* adalah nilai yang menandakan minggu ke berapa sebuah rekening dibuka. Nilai *week_open* ini berkisar antara 1-52.

5.3.3 Atribut *tenuredays, tenureweek, tenuremonth*

Atribut *tenuredays, tenureweek, tenuremonth* merupakan atribut hasil proses agregasi dari dua buah variabel yaitu, atribut *date open* dan atribut *date close*. Seperti yang telah dijelaskan pada bab sebelumnya *tenure* sebuah adalah lama relationship antara customer dan perusahaan. *Tenure* dalam tesis ini adalah lama hidup sebuah account yang dihitung berdasarkan selisih tanggal ketika account tersebut dibuka (start date) dan tanggal ketika account tersebut ditutup. Atribut *tenuredays, tenureweek, tenuremonth* juga memakai konsep yang sama dengan pembedaan pada satuan yang dipakai dalam menghitung *tenure*. Atribut *tenuredays* menggunakan satuan hari, *tenureweek* menggunakan satuan minggu, sedangkan *tenuremonth* menggunakan satuan bulan.

5.3.4 Tabel *balance*

Tabel balance merupakan tabel yang menyimpan data historis mengenai fluktuasi saldo harian dari nasabah. Tabel ini dibentuk dari atribut *balance date* yang ada pada *dataset* awal. Kolom yang terbentuk dalam tabel ini bisa dilihat pada tabel 4 dibawah ini.

Nama Kolom	Keterangan
Regno	*FK*
Date	Tanggal saldo harian
Balance	Saldo Harian

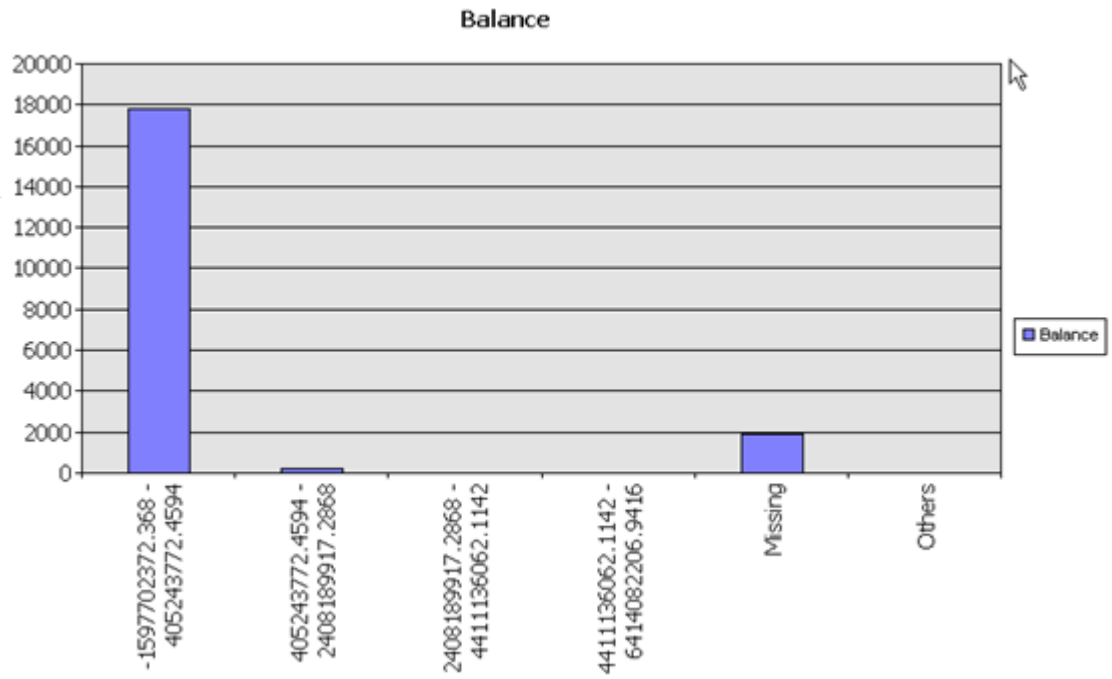
Tabel 4. Tabel balance dan keteranganya

Data historis yang tersimpan pada tabel balance berjumlah cukup besar yaitu sebanyak 1308013 baris. Besarnya data yang ada dapat menimbulkan permasalahan berupa lambatnya peforma proses *mining* nantinya. Untuk itu perlu dilakukan pengurangan data melalui agregasi data. Proses agregasi dilakukan dengan mengubah nilai saldo harian menjadi nilai rata-rata saldo mingguan.

Proses agregasi data berhasil mengurangi jumlah data secara signifikan, dari 1308013 baris data menjadi 280853 baris data. Hasil reduksi data tersebut disimpan kedalam sebuah tabel baru bernama *balance_weekly_avg*.

5.3.5 Diskritisasi

Hasil agregasi data yang tersimpan pada tabel *balance_weekly_avg* menghasilkan variasi nilai atribut *balance* yang sangat jauh yaitu dari -3,802,791,238.885 sampai ke 49,197,364,713.48. Untuk melihat jenis sebaran data maka dilakukan *sampling* secara acak sebanyak 20000 sample, lalu dikelompokkan ke dalam empat kelompok dengan interval nilai yang sama (*equal-widht discretization*). Hasil dari sampling tersebut terlihat pada gambar 15 dibawah. Dari gambar tersebut terlihat bahwa data mengumpul pada sebuah interval tertentu atau dengan kata lain data tidak tersebar secara merata. Untuk mencegah hal tersebut dan untuk mempermudah intepretasi hasil mining nanti, maka perlu dilakukan diskritisasi secara *equal-depth*. Teknik diskritisasi secara equal-depth ini penulis adopsi dari [15].



Gambar 15. Sampling acak field balance dari tabel balance_week_avg

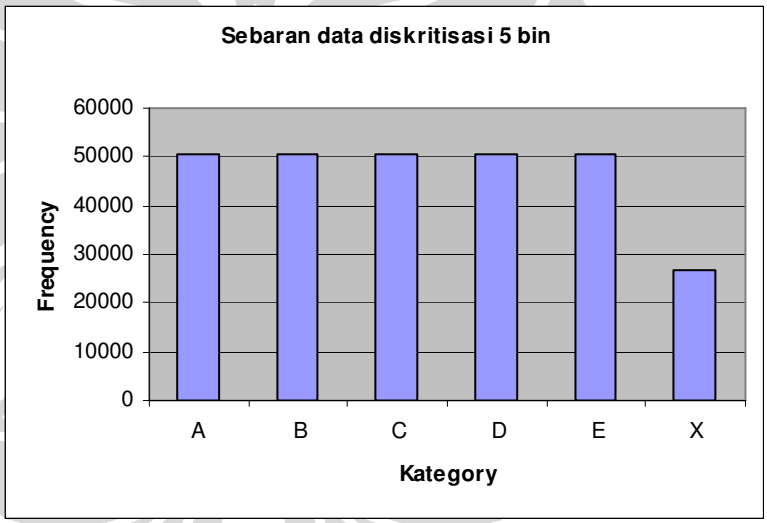
Pada thesis ini jumlah bin yang dipakai dibagi kedalam 3 jenis, yaitu 5, 7, dan 9 bin. Tujuan dari hal ini adalah untuk memperoleh perbandingan tingkat akurasi model yang akan terbentuk nanti. Ketika melakukan diskritisasi, penulis mendapati terdapat beberapa nilai yang hilang (*null value*) dari *dataset* yang akan diolah. Nilai *null* ini merupakan sebuah penanda bahwa seorang nasabah sudah menutup rekeningnya, karena itu nilai ini akan dimasukkan kedalam sebuah *bin* tambahan tersendiri.

Diskritisasi dilakukan terhadap field *balance* dari tabel *balance_weekly_avg* tersebut hasilnya disimpan kedalam 3 buah field baru, yaitu binning1, binning2, dan binning3. Binning1 menyimpan hasil diskritisasi berjenis 5, binning2 menyimpan hasil disktritisasi berjenis 7, dan binning 9 menyimpan diskritisasi berjenis 9.

Diskritisasi jenis 5 membagi nilai field *balance* kedalam 6 kategori yaitu A, B, C, D, E, dan X. Kategori X merupakan kategori khusus yang mengkategorikan nilai *null* yang terdapat pada field *balance*. Sebaran dari hasil diskritisasi ini bisa dilihat pada tabel 5 dan gambar 16 dibawah ini.

Kategori	Rentang nilai field balance
A	<83152.19
B	>=83152.19 AND <911828.66
C	>=911828.66 AND <3673132
D	>=3673132 AND <16332204.64
E	>=16332204.64
X	*null*

Tabel 5. Kategorisasi diskritisasi jenis 5

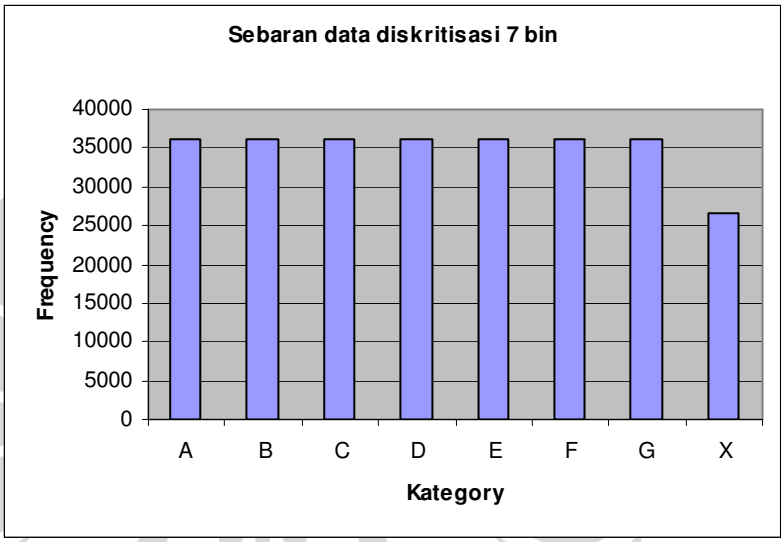


Gambar 16. Sebaran data diskritisasi jenis 5

Sama seperti diskritisasi jenis 5, diskritisasi jenis 7 membagi nilai field *balance* kedalam 8 kategori yaitu A, B, C, D, E, F, G dan X. Serupa seperti diskritisasi yang lain, kategori X merupakan kategori khusus yang mengkategorikan nilai *null* yang terdapat pada field *balance*. Sebaran dari hasil diskritisasi ini bisa dilihat pada tabel 6 dan gambar 17 dibawah ini.

Kategori	Rentang nilai field balance
A	< 39859
B	>= 39859 AND < 300710
C	>= 300710 AND < 1136812.32
D	>= 1136812.32 AND < 3154718
E	>= 3154718 AND < 8097714
F	>= 8097714 AND < 30715118.25
G	>= 30715118.25
X	*null*

Tabel 6. Kategorisasi diskritisasi jenis 7

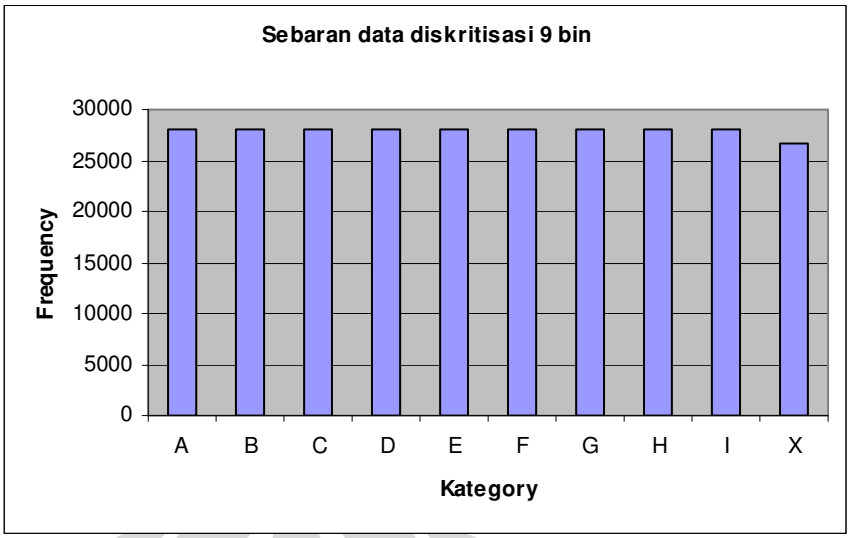


Gambar 17. Sebaran data diskritisasi jenis 7

Kategori diskrit untuk diskritisasi jenis 9, terbagi kedalam 10 kategori, yaitu A, B, C,...,I dan X untuk kategori nilai null. Sebaran dan hasil diskritisasi dapat dilihat pada tabel 7 dan gambar 18.

Kategori	Rentang nilai field balance
A	< 20002
B	>= 20002 AND < 115501.29
C	>= 115501.29 AND < 490136.83
D	>= 490136.83 AND < 1282497
E	>= 1282497 AND < 2894242.5
F	>= 2894242.5 AND < 5882896.41
G	>= 5882896.41 AND < 13270537
H	>= 13270537AND < 46288517.193
I	>= 46288517.193
X	*null*

Tabel 7. Kategorisasi diskritisasi jenis 9



Gambar 18. Sebaran data diskritisasi jenis 9

5.4 Pembentukan tabel yang akan di proses

Setelah proses persiapan data selesai dilakukan, maka proses dilakukan proses mapping ke dalam tabel-tabel baru yang akan siap diolah dengan memakai software *data mining*. Tabel-tabel tersebut antara lain.

5.4.1 Tabel Acc_info

Seperti telah dijelaskan pada sub bab sebelumnya, tabel acc_info secara umum merupakan kumpulan variabel-atribut *dataset* awal tanpa menyertakan atribut balance. Struktur dari tabel acc_info dapat dilihat pada tabel 8 dibawah ini.

Column Name	Type	Size
REGNO	Long Integer (PK)	4
BRANCH CODE	Text	4
SEX	Text	8
PRODUCT	Text	32
CUSTTYPE	Text	32
CUST CLASS	Text	16
DATI II	Text	64
PROVINCE	Text	64
HasATMCard	Yes/No	1
ReferalBy	Text	50
JoinDate	Date/Time	8
dtopen	Double	8
mtopen	Double	8
yropen	Double	8
OPEN DATE	Date/Time	8
CLOSE DATE	Date/Time	8
week close	Byte	1
week open	Byte	1
tenuredays	Long Integer	4
tenureweek	Long Integer	4
tenuremonth	Long Integer	4
Close_status	Yes/No	1

Tabel 8. Struktur tabel acc_info

5.4.2 Tabel Balance_weekly_avg

Tabel balance_weekly_avg merupakan tabel historis fluktuasi saldo mingguan rata-rata yang merupakan hasil agregasi dan diskritisasi dari tabel balance. Struktur dari tabel *balance_weekly_avg* dapat dilihat pada tabel dibawah ini.

Column Name	Type	Size
id	Long Integer	4
Regno	Long Integer	4
week	Long Integer	4
Balance	Currency	8
Binning1	Text	1
Binning2	Text	1
Binning3	Text	1

Tabel 9. Struktur tabel balance_weekly_avg

BAB VI
PEMODELAN DAN EVALUASI

Data yang sudah dipersiapkan sebelumnya akan diolah dengan teknik *data mining* untuk mendapatkan model *data mining* sehiingga bisa dievaluasi untuk mengangkat informasi tersembunyi yang terkandung dalam kumpulan data yang ada. Bab ini akan membahas mengenai pemakaian teknik *data mining* terhadap data yang sudah siap dan hasil evaluasi terhadap model yang dihasilkan.

Sebelum melangkah lebih jauh ke dalam pemodelan *data mining* perlu diperjelas apa yang ingin dihasilkan dari *data mining* itu sendiri. Hasil yang diharapkan dari *data mining* tentu adalah jawaban dari pertanyaan masalah yang dihadapi. Pertanyaan-pertanyaan tersebut adalah:

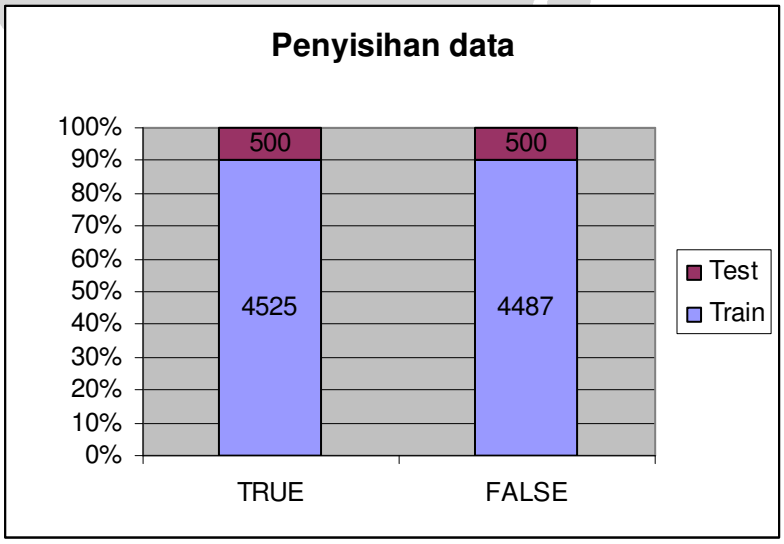
- "Nasabah-nasabah yang seperti apakah yang berpotensi akan menutup rekening tabungan mereka?"
- "*Pattern* yang seperti apa yang bisa digunakan untuk mendeteksi nasabah-nasabah yang akan menutup rekening tersebut?"

Untuk dapat menjawab pertanyaan diatas maka model *data mining* yang digunakan dalam thesis ini adalah model yang menggunakan teknik *Microsoft Sequence clustering*. Alasan utama dipilihnya teknik ini adalah kemampuan teknik ini untuk mengelompokan gabungan *variabel sequence* dan *non-sequence* kedalam segmen-segmen yang lebih homogen berdasarkan kemiripannya dengan pola (*pattern*) sequence yang ada. Algoritma *Microsoft Sequence clustering* itu sendiri merupakan salah satu algoritma yang tersedia dalam paket perangkat lunak *database Microsoft SQL Server 2005.*

6.1 Pembentukan model

Model *sequence clustering* dibentuk dengan menggunakan data yang tersimpan dalam tabel *acc_info* dan tabel *balance_avg_weekly*. Tabel acc_info dipakai sebagai *case* tabel, sedangkan tabel *balance_avg_weekly* digunakan sebagai *nested* tabel.

Dari seluruh data yang terdapat pada tabel acc_info tersebut tidak seluruhnya digunakan dalam proses pemodelan. Total 1000 data dari keseluruhan data dipisahkan untuk tujuan evaluasi akurasi model (*test dataset*) sedangkan sisanya dijadikan sebagai dataset pembentuk model (*train dataset*). Pemisahan data dilakukan secara random berimbang berdasarkan kategori *close_status*, dimana sebanyak 500 *record* diambil secara random dari rekening-rekening yang sudah tutup (*close_status = true*), dan 500 *record* lainnya diambil dari rekening-rekening yang masih aktif (*close_status = false*). Gambar 19 dibawah menggambarkan lebih jelas proporsi dari *train dataset* dan *test dataset*.



Gambar 19. Grafik distribusi pembagian dataset train dan test

6.2 Pemilihan atribut

Variabel yang dipakai dalam pembentukan model ditentukan oleh tingkat korelasi variabel tersebut dengan variabel target, dalam hal ini variabel target adalah variabel *close_status*. Untuk menentukan tingkat korelasi variabel tersebut dengan variabel target, penulis memakai fungsi *feature selection* yang telah disediakan oleh *SQL Server 2005*. Fungsi ini dapat melakukan analisa singkat untuk memilih variabel mana saja yang memiliki tingkat keterkaitan dengan variabel target.

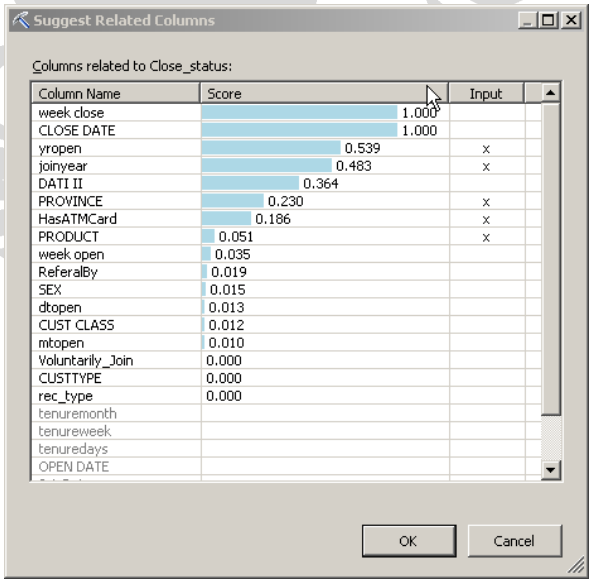
Gambar 20. dibawah merupakan tampilan dari fungsi tersebut. Dari gambar tersebut dapat dilihat tingkat keterkaitan antara variabel-variabel lain dengan variabel *close_status*. Tingkat keterkaitan dinilai melalui kolom *score*, semakin besar korelasi sebuah variabel, semakin besar nilai *score* yang dihasilkan. Nilai skor tertinggi adalah 1.0 sedangkan terendah adalah 0. Pemilihan variabel yang akan dipakai dalam proses pembentukan model nanti didasarkan pada skor korelasi yang dimiliki oleh variabel tersebut terhadap variabel *close_status*. Batasan nilai minimum skor yang dipakai penulis dalam memilih variabel adalah 0,05.

Dengan memakai batasan tersebut didapatkan beberapa kandidat variabel yang akan dipilih, yaitu: *week close, close date, yropen, joinyear, dati ii, province, HasATMCard, Product.* Dari kandidat-kandidat tersebut tidak seluruhnya digunakan dalam proses pembentukan model. Terlihat bahwa variabel *week close, close date* adalah variabel kandidat dengan skor tertinggi. Akan tetapi ketika ditelaah lebih dalam, variabel tersebut merupakan sebuah *false predictor* [12], sehingga menjadikan variabel tersebut sebaiknya tidak digunakan dalam proses

pembentukan model. Variabel *DATI II* walaupun memiliki korelasi cukup besar dengan variabel target, tidak diikutsertakan dalam proses selanjutnya, hal ini dikarenakan jumlah variasi nilai yang cukup besar yang dimiliki oleh variabel tersebut (118 variasi) selain itu variabel tersebut juga sudah terwakili oleh variabel *Province*.

Variabel-variabel yang memiliki skor dibawah 0,05 seperti variabel *Sex, Custclass, Custtype, Voluntairly_Join*, walaupun secara intuitif sangat menjanjikan untuk digunakan sebagai variabel input, namun dikarenakan rendahnya skor yang dimiliki oleh variabel-variabel tersebut jika dipaksa digunakan, dapat membuat pola-pola yang didapat nantinya menjadi tidak begitu berarti (*meaningless*).

Setelah dilakukan pemilahan kandidat variabel maka didapat variabel-variabel yang akan digunakan dalam proses pembentukan model, variabel-variabel tersebut adalah*, yropen, joinyear, province, HasATMCard, Product.*



Gambar 20. Pemilihan Variabel

Jumlah model yang dibuat dengan memakai teknik *sequence clustering* ini berjumlah 3 buah model. Ketiga buah model tersebut pada dasarnya adalah model

yang memakai *input* yang sama, yaitu dataset *train*, yang membedakan ketiga buah model tersebut adalah jenis variabel diskritisasi yang tersimpan pada tabel tabel *balance_weekly_avg*. Ketiga model itu seperti yang terlihat juga pada tabel 6-1 dan 6-2 dibawah adalah:

- Model binning5, yang menggunakan variabel diskritisasi jenis 5 (*field* binning1).
- Model binning7, yang menggunakan variabel diskritisasi jenis 7 (*field* binning2), dan
- Model binning9, yang menggunakan variabel diskritisasi jenis 9 (*field* binning3).

	Model binning5			Model binning7			Model binning9		
Nama Atribut	Key	input	predict	Key	input	predict	Key	input	predict
Close Status		√	√		√	√		√	√
Product		√			√			√	
Province		√			√			√	
Regno	√			√			√		
HasATMCard		√			√			√	

Tabel 10. Perbandingan atribut *case* tabel dari ketiga model

	Model binning5			Model binning7			Model binning9		
Nama Atribut	FK	Key Sequence	predict	FK	Key Sequence	predict	FK	Key Sequence	predict
Regno	√			√			√		
week		√			√			√	
Binning1			√						
Binning2						√			
Binning3									√

Tabel 11. Perbandingan atribut *nested* tabel dari ketiga model

6.3 Evaluasi model

Model yang sudah dibentuk tentu butuh untuk diuji keakuratannya. Dari ketiga model tersebut nantinya akan dipilih salah satu model dengan tingkat akurasi model yang paling tinggi. Model yang terpilih kemudian dianalisa untuk agar didapatkan informasi berharga. Metode yang dipakai untuk mengukur tingkat akurasi model adalah metode matrik klasifikasi. Variabel dari *case* tabel yang dipakai adalah variabel *close_status*, yang memiliki dua buah kelas, yaitu *true* dan *false*

6.3.1 Matrik klasifikasi

Matrik klasifikasi yang sering juga disebut *confusion matrik* merupakan sebuah matrik yang menunjukkan berapa kali sebuah algoritma memprediksi hasil secara tepat, metode ini dipakai penulis untuk memilih model mana yang terbaik untuk dianalisa lebih lanjut. Matrik klasifikasi juga dapat dipakai untuk mengukur *error rate* atau peforma sebuah model.

Kriteria yang digunakan penulis untuk memilih model yang terbaik adalah dengan memakai metode *10-fold cross validation* (FCV) seperti yang disarankan oleh [14]. Metode FCV ini membagi dataset kedalam 10 bagian yang berbeda secara acak. Tiap bagian akan digunakan sekali sebagai *test dataset* dan sisanya digunakan sebagai *train dataset*. Lalu model dijalankan dengan memakai *train dataset* yang dipilih, hingga didapatkan tingkat akurasi dan *error rate*. Proses ini diulang sampai 10 kali dengan memakai bagian yang lain. Kesepuluh tingkat akurasi dan *error rate* kemudian dirata-rata. Model yang terpilih nantinya untuk dianalisa lebih lanjut adalah model dengan tingkat akurasi rata-rata tertinggi dan *error rate* rata-rata terendah. Tabel 6-1, tabel 6-2, dan tabel 6-3 dibawah

menunjukkan perbandingan matrik klasifikasi dan pengukuran peforma dengan memakai teknik *10-fold cross validation* antara model binning5, binning7, dan model binning9.

Dari ketiga tabel tersebut terlihat bahwa model binning5 memiliki persentase tertinggi untuk tingkat akurasi rata-rata. Model binning7 juga memiliki persentase terendah dalam parameter *error rate* rata-rata. Oleh karena itu model binning5 inilah yang nantinya dipakai dalam analisa lebih lanjut.

Bin5	Predicted	True (Actual)	False (Actual)	True Positive	True Negatif	False positif	False Negative	Akurasi (AC)	ER = 1 - AC
RUN1	TRUE	479	6	95.80%	98.80%	4.08%	4.20%	97.30%	2.70%
	FALSE	21	494						
RUN2	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN3	TRUE	500	396	100.00%	20.80%	0.00%	0.00%	60.40%	39.60%
	FALSE	0	104						
RUN4	TRUE	464	6	92.80%	98.80%	6.79%	7.20%	95.80%	4.20%
	FALSE	36	494						
RUN5	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN6	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN7	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN8	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN9	TRUE	499	393	99.80%	21.40%	0.93%	0.20%	60.60%	39.40%
	FALSE	1	107						
RUN10	TRUE	488	214	92.95%	56.06%	11.94%	7.05%	75.20%	24.80%
	FALSE	37	273						
				Average error rate (true error rate)					36.07%
				Average accuracy					63.93%

NA = Nilai yang tidak bisa dihitung karean bilangan pembaginya adalah 0

Tabel 12. 10 Fold Cross Validation model binning 5

Bin7	Predicted	True (Actual)	False (Actual)	True Positive	True Negatif	False positif	False Negative	Akurasi (AC)	ER = 1 - AC
RUN1	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN2	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN3	TRUE	479	338	95.80%	32.40%	11.48%	4.20%	64.10%	35.90%
	FALSE	21	162						
RUN4	TRUE	492	290	98.40%	42.00%	3.67%	1.60%	70.20%	29.80%
	FALSE	8	210						
RUN5	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN6	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN7	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN8	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN9	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
	FALSE	0	0						
RUN10	TRUE	479	121	91.24%	75.15%	11.17%	8.76%	83.50%	16.50%
	FALSE	46	366						
				Average error rate (true error rate)					43.22%
				Average accuracy					56.78%

NA = Nilai yang tidak bisa dihitung karean bilangan pembaginya adalah 0

Tabel 13. 10 Fold Cross Validation model binning 7

Bin9	Predicted	True (Actual)	False (Actual)	True Positive	True Negatif	False positif	False Negative	Akurasi (AC)	ER = 1 - AC
RUN1	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
seed = 8198	FALSE	0	0						
RUN2	TRUE	489	301	97.80%	39.80%	5.24%	2.20%	68.80%	31.20%
seed = 3475	FALSE	11	199						
RUN3	TRUE	479	316	95.80%	36.80%	10.24%	4.20%	66.30%	33.70%
seed = 5969	FALSE	21	184						
RUN4	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
seed = 2841	FALSE	0	0						
RUN5	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
seed = 7402	FALSE	0	0						
RUN6	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
seed = 4015	FALSE	0	0						
RUN7	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
seed = 994	FALSE	0	0						
RUN8	TRUE	500	500	100.00%	0.00%	NA	0.00%	50.00%	50.00%
seed = 6354	FALSE	0	0						
RUN9	TRUE	500	282	100.00%	43.60%	0.00%	0.00%	71.80%	28.20%
seed = 1589	FALSE	0	218						
RUN10	TRUE	525	487	100.00%	0.00%	NA	0.00%	51.88%	48.12%
seed = 9424	FALSE	0	0						
				Average error rate (true error rate)					44.12%
				Average accuracy					55.88%

NA = Nilai yang tidak bisa dihitung karean bilangan pembaginya adalah 0

Tabel 14. 10 Fold Cross Validation model binning 9

BAB VII
EMAS YANG DIPEROLEH

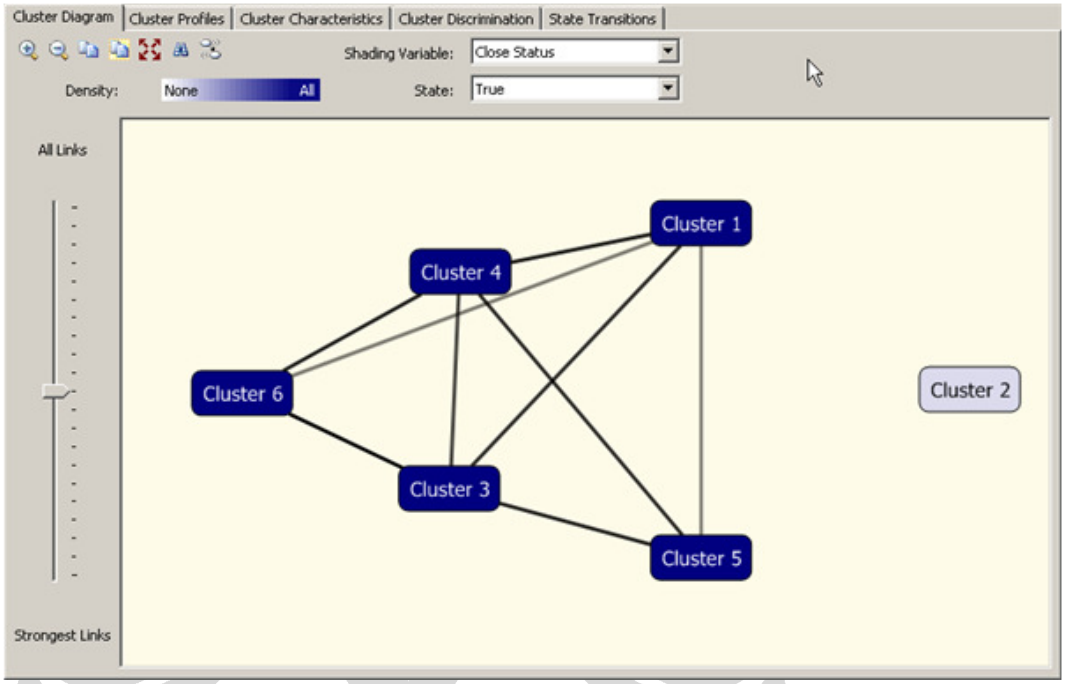
Seperti telah dikemukakan pada bab sebelumnya, tujuan dari *data mining* adalah untuk memperoleh informasi berharga dari model yang dibuat. Informasi tersebut didapat melalui analisa lebih lanjut atas pola-pola yang ditemukan dari model yang dibuat. Berdasarkan informasi yang didapat tersebut pula kemudian ditarik kesimpulan umum.

7.1 Analisa lebih lanjut model binning5

Model terbaik yang dihasilkan dari evaluasi model adalah model binning. Model ini berhasil mengelompokkan pola-pola transaksi mingguan dari nasabah kedalam 6 buah kelompok yang berbeda. Dari diagram kluster seperti yang terlihat pada gambar 21 dibawah, terlihat bahwa kluster-kluster yang memiliki kemiripan yang sama saling berkumpul satu dengan yang lain. Dari gambar tersebut terlihat bahwa kluster 2 merupakan kluster tersendiri yang memisah dari kumpulan kluster yang lain.

Tingkat kepadatan sebuah kluster terkait dengan nilai *variabel* tertentu digambarkan melalui gradasi warna latar belakang kluster tersebut, semakin gelap warna latar belakang kluster tersebut, semakin besar tingkat kepadatannya. Gambar 21 menggambarkan juga densitas kluster-kluster yang mengelompokkan rekening-rekening nasabah yang sudah tutup dalam hal ini direpresentasikan oleh nilai *variabel close_status=true*. Dari gambar tersebut terlihat juga bahwa kluster 1, 3, 4, 5, dan 6 adalah kluster-kluster dengan tingkat densitas yang tinggi untuk rekening-rekening nasabah yang sudah ditutup. Sedangkan kluster 2 adalah kluster

dari kumpulan rekening-rekening nasabah yang belum ditutup, dalam hal ini direpresentasikan oleh nilai variabel *close_status* = *false*.



Gambar 21. Diagram kluster model binning5

Algoritma Microsoft sequential clustering menggunakan algoritma EM (*Expectation Maximization*) untuk melakukan pengklusteran. Karena Algoritma EM *clustering* termasuk kedalam salah satu jenis algoritma *soft clustering*, maka cara algoritma ini menempatkan sebuah objek terhadap sebuah kluster saling tumpang tindih berdasarkan nilai kemungkinan objek tersebut berada dalam kluster tertentu, dengan begitu sebuah objek bisa berada dalam lebih dari satu kluster[10]. Hal ini mengakibatkan kluster-kluster yang dihasilkan oleh model binning5 memiliki tingkat kemiripan antar kluster yang cukup tinggi. Gambaran karakteristik dari kluster-kluster yang dihasilkan oleh model binning5 adalah dideskripsikan dalam beberapa sub bab berikut.

7.1.1 Cluster 1

Cluster 1 merupakan salah satu kluster dari kumpulan rekening-rekening dengan status sudah ditutup (*close_status = true*). Tidak ada rekening dari kluster ini yang masih berstatus aktif (*close_status = false*). Sebagian besar pemilik rekening dalam kluster ini adalah nasabah yang tidak menggunakan atau tidak memakai fasilitas kartu ATM (*Has ATM Card = false*). Selain itu, sebagian besar rekening yang berada di dalam *cluster 1* adalah rekening dari produk tabungan (PRODUCT =Tabungan). Jika dilihat dari sisi demografis tempat nasabah berada, nasabah-nasabah pemilik rekening yang berada dalam C*luster 1*, memiliki kecenderungan berasal dari provinsi DKI Jakarta.

Nasabah pemilik rekening yang pertama kali bergabung diatas tahun 2002 (*Join Year >= 2002*) atau bergabung antara tahun 1998 sampai tahun 2002 (*Join Year = 1998 – 2002*) memiliki kecenderungan cukup besar berada dalam kluster ini. Selain itu, rekening-rekening yang dibuka diatas tahun 2003 (Yropen >= 2003) juga memiliki tingkat propabilitas yang cukup tinggi untuk berada di kluster ini. Tabel karakteristik yang lengkap dari kluster ini bisa dilihat pada lampiran 1.

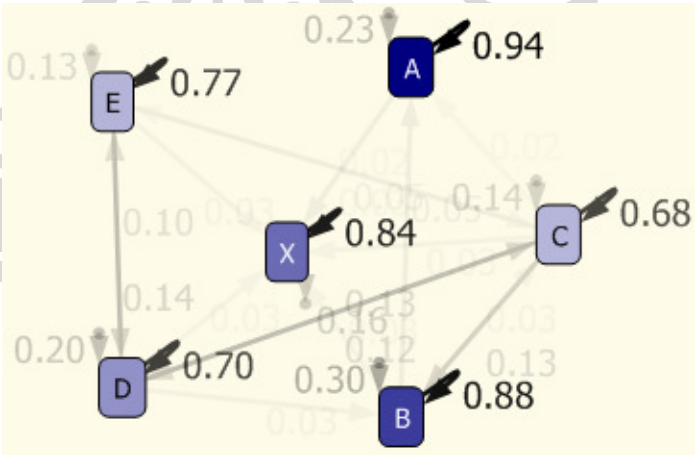
Gambar 22, menggambarkan diagram transisi keadaan dari *cluster 1*. Informasi diagram transisi keadaan tersebut dibuat berdasarkan matrik transisi keadaan yang dapat dilihat pada tabel 15. Matrik tersebut hanya menunjukkan transisi keadaan yang memiliki tingkat propabilitas lebih besar dari nol[10].

Melalui matrik dan diagram transisi keadaan tersebut dapat kita lihat bahwa rekening-rekening yang berada pada *cluster 1*, sebagian besar rekening mingguannya sebagian besar berada pada level A. Dari diagram dan tabel tersebut juga ditemukan bahwa sebagian besar rekening memiliki kecenderungan untuk

berada pada level yang sama di minggu yang akan datang. Hal ini terlihat dari besarnya propabilitas transisi perpindahan untuk level yang sama diatas angka 0.65. Akan tetapi kita juga bisa melihat ada sedikit kemungkinan rekening-rekening pada kluster ini untuk menutup rekeningnya minggu depan. Sebagai contoh dari dari tabel 15 terlihat, jika sebuah rekening pada minggu ini memiliki saldo rekening rata-rata berada pada level C, maka terdapat sekitar 3% kemungkinan rekening tersebut akan tutup minggu depan (sel yang menandakan angka ini berada pada baris ke 3 kolom ke 6).

	A	B	C	D	E	X
A	0.94					0.05
B	0.05	0.88	0.03	0.03		
C	0.02	0.13	0.68	0.12	0.02	0.03
D		0.03	0.13	0.70	0.10	0.03
E			0.03	0.14	0.77	0.03
X						0.84

Tabel 15. Matrik transisi keadaan cluster 1



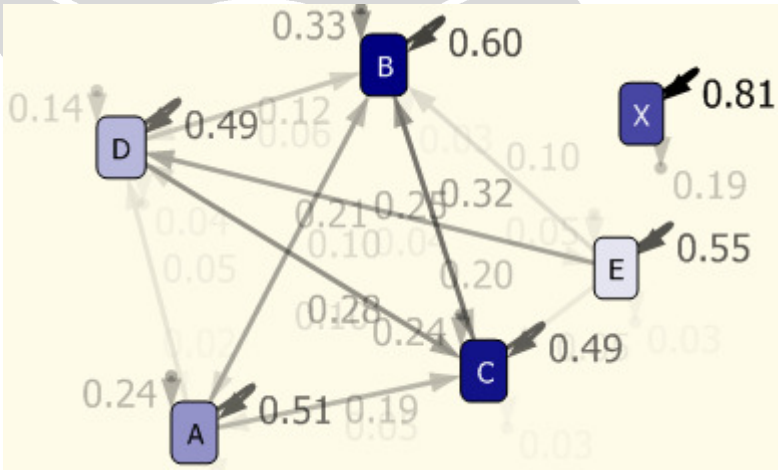
Gambar 22. Diagram transisi keadaan cluster 1

7.1.2 Cluster 2

Nasabah-nasabah yang memiliki rekening-rekening pada *cluster 2* memiliki kecenderungan sangat besar untuk tidak menutup rekening mereka (*close_status = false*). Sebagian besar rekening yang berada pada kluster ini adalah rekening dari

produk tabungan. Perbandingan propabilitas antara nasabah yang memakai kartu ATM dan nasabah yang tidak memakai kartu atm hampir sama. Dari sisi demografis, nasabah yang berasal dari daerah Jawa Barat atau DKI jakarta memiliki kecenderungan untuk berada pada cluster ini. Tabel karakteristik selengkapnya dari kluster ini bisa dilihat pada lampiran 2.

Dari matrik dan diagram transisi keadaan *cluster 2* seperti yang terlihat pada gambar 23 dan tabel 16, terlihat bahwa sebagian besar rekening yang berada pada kluster ini adalah rekening-rekening yang memiliki saldo rata-rata mingguan pada level B dan C. Dari matrik tersebut terlihat juga bahwa rekening-rekening pada kluster ini memiliki fluktuasi saldo rata-rata mingguan yang cukup dinamis, hal ini terlihat dari angka propabilitas perpindahan antar level yang cukup besar, sementara propabilitas untuk tetap berada di level yang sama terlihat lebih kecil dibandingkan dengan propabilitas yang terdapat di kluster yang lain. Dari matrik transisi tersebut dapat juga terlihat bahwa nasabah-nasabah yang berada di cluster ini tidak akan menutup rekening mereka.



Gambar 23. Diagram transisi keadaan cluster 2

	A	B	C	D	E	X
A	0.51	0.21	0.19	0.05		
B	0.10	0.60	0.20	0.06		
C	0.05	0.32	0.49	0.10		
D	0.02	0.12	0.28	0.49	0.04	
E		0.10	0.05	0.25	0.55	
X						0.81

Tabel 16. Matrik transisi keadaan cluster 2

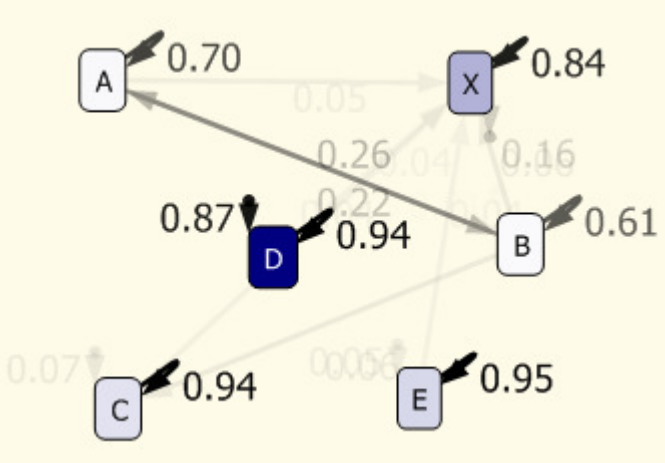
7.1.3 Cluster 3

Cluster 3 adalah salah satu kluster yang memiliki populasi terbanyak kedua dari model binning5. Seluruh rekening-rekening yang diklasifikasikan ke dalam kluster ini adalah rekening-rekening yang sudah tutup (*close_status = true*). Dari tabel karakteristik yang ada pada lampiran 3 terlihat juga bahwa sebagian besar rekening yang masuk kedalam kluster ini nasabahnya tidak memiliki kartu ATM (*Has ATM Card = false*). Jika dilihat dari sisi demografis nasabah, ada kecenderungan cukup besar nasabah dari provinsi DKI Jakarta terklasifikasi kedalam cluster ini. Karakteristik lengkap dari kluster ini bisa dilihat pada lampiran 3.

Fluktuasi saldo rata-rata mingguan dari *cluster 3* ini dapat dianalisa melalui diagram dan matrik transisi keadaan seperti yang terlihat pada gambar 24 dan tabel 17 dibawah. Dari diagram transisi keadaan tersebut terlihat bahwa sebagian besar rekening yang berada di kluster ini adalah rekening-rekening yang memiliki fluktuasi saldo mingguan rata-rata di level D. Dari tabel transisi keadaan juga terlihat bahwa rekening-rekening di kluster ini cenderung berfluktuasi secara statis, dalam arti propabilitas perpindahan antar level memiliki nilai cukup kecil, sedangkan propabilitas untuk tetap berada pada level yang sama cukup besar.

Dari tabel transisi tersebut terlihat juga bahwa rekening-rekening pada kluster ini memiliki kecenderungan untuk menutup rekeningnya di minggu yang

akan datang, hal ini terlihat dari adanya nilai propabilitas transisi dari tiap level saldo mingguan rata-rata ke level X.



Gambar 24. Diagram Transisi keadaan cluster 3

	A	B	C	D	E	X
A	0.70	0.22				0.05
B	0.26	0.61	0.06			0.06
C			0.94			0.04
D				0.94		0.04
E					0.95	0.09
X						0.84

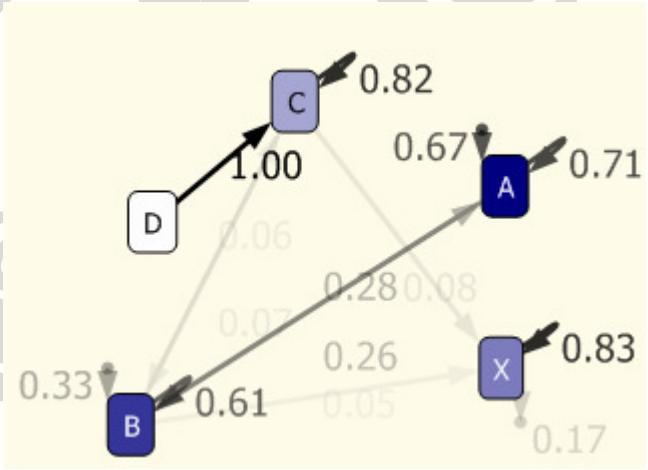
Tabel 17. Matrik transisi keadaan cluster 3

7.1.4 Cluster 4

Rekening-rekening yang berada di *cluster 4* hampir dipastikan adalah rekening-rekening yang memiliki status sudah tutup (*Close_status = true*). Dari tabel karakeristik *cluster 4* yang ada pada lembar lampiran 4, terlihat juga bahwa nasabah pemilik rekening-rekening yang berada di kluster ini sangat besar kemungkinannya tidak memakai fasilitas kartu ATM (*Has ATM Card = false*). Selain itu terdapat juga kecenderungan besar bahwa nasabah yang memiliki rekening di kluster ini adalah nasabah-nasabah yang pertama kali bergabung di rentang tahun 1996-1998 (*Joinyear = 1996-1998*), sedangkan rekening-rekening

yang dibuka antara tahun 1997 – 1999 (*Yropen = 1997-1999*) atau lebih memiliki kecenderungan besar untuk berada di dalam kluster ini.

Diagram dan matrik transisi keadaan dari kluster ini terlihat pada gambar 25 dan tabel 18 dibawah memperlihatkan fluktuasi dari saldo mingguan rekening-rekening yang tergabung dalam kluster ini. Dari gambar dan tabel tersebut terlihat bahwa sebagian besar fluktuasi saldo mingguan rata-rata dari rekening yang tergabung dalam *cluster 4* adalah rekening-rekening yang jumlah saldo mingguan rata-ratanya berada pada level A. Dari gambar dan matrik tersebut juga ditemukan fluktuasi perpindahan antar level yang cukup kecil, dengan kata lain bisa dikatakan bahwa fluktuasi saldo mingguan rata-rata dari rekening-rekening yang terdapat di kluster ini cenderung statis.



Gambar 25. Diagram Transisi keadaan cluster 4

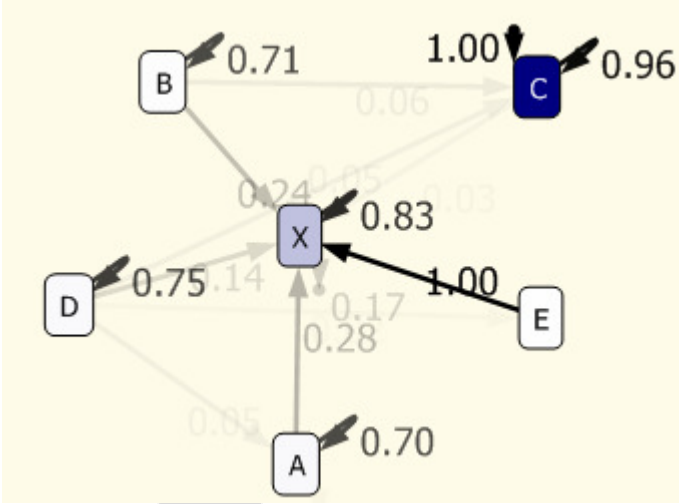
	A	B	C	D	E	X
A	0.71	0.26				
B	0.28	0.61	0.06			0.05
C		0.07	0.82			0.08
D			1.00			
E						
X						0.83

Tabel 18. Matrik transisi keadaan cluster 4

7.1.5 Cluster 5

Cluster 5 merupakan salah satu kluster tempat berkumpulnya rekening-rekening dengan status sudah tutup. Nasabah-nasabah yang bergabung antara tahun 1996-1998 (*Joinyear = 1996-1998*) memiliki kecenderungan cukup besar untuk bergabung ke dalam kluster ini. Propabilitas nasabah yang tidak memakai fasilitas kartu ATM untuk tergabung ke dalam kluster ini cukup besar. Tabel karakteristik *cluster 5* yang terdapat pada lembar lampiran juga menunjukkan bahwa, lebih dari separuh rekening yang berada pada kluster ini adalah rekening-rekening yang dibuka antara tahun 1997 sampai tahun 1999 (*Yropen = 1997-1999*). Nasabah-nasabah yang memakai produk tabungan berjangka (*Product = Tabungan Berjangka*) memiliki kemungkinan cukup besar untuk tergabung ke dalam kluster ini.

Diagram dan matrik transisi dari kluster ini dapat dilihat pada gambar 26 dan tabel 19 dibawah. Dari gambar tersebut terlihat bahwa sebagian besar fluktuasi saldo rata-rata mingguan dari rekening-rekening yang tergabung dalam kluster ini berada pada level C. Dari matrik transisi tersebut juga ditemukan kecenderungan yang cukup besar dari nasabah dengan rekening-rekening yang berada pada level A, B, atau D pada minggu ini, untuk menutup rekening mereka di minggu yang akan datang. Dari kedua diagram dan matrik tersebut juga terlihat bahwa rekening-rekening yang berada dalam *cluster 5* memiliki kecenderungan besar untuk bergerak statis pada level yang sama.



Gambar 26. Diagram Transisi keadaan cluster 5

	A	B	C	D	E	X
A	0.70					0.28
B		0.71	0.06			0.24
C			0.96			0.03
D	0.05		0.05	0.75	0.02	0.14
E						1.00
X						0.83

Tabel 19. Matrik transisi keadaan cluster 5

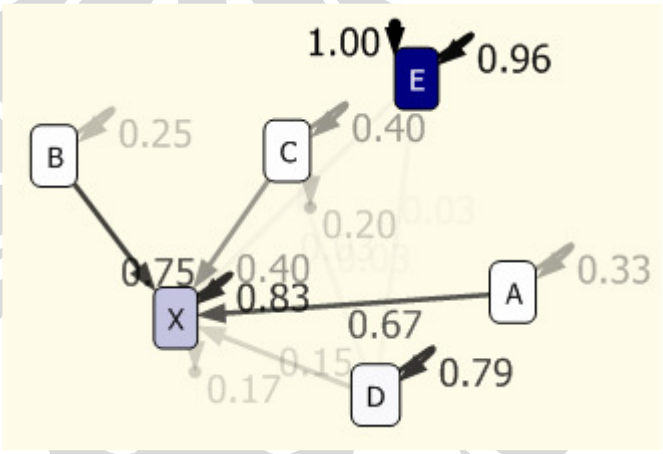
7.1.6 Cluster 6

Nasabah-nasabah yang rekeningnya berada pada kluster ini adalah nasabah-nasabah yang sudah menutup rekeningnya. Kecenderungan penggunaan fasilitas kartu ATM dari nasabah yang rekeningnya tergabung dalam kluster ini cukup kecil, selain itu nasabah yang bergabung antara tahun 1993 - 2002 (*Joinyear = 1993-1996, 1996-1998, 1998-2002*) memiliki kecenderungan besar untuk berada di dalam kluster ini. Rekening-rekening yang dibuka pada tahun 1997 - 1999 juga memiliki kecenderungan untuk tergabung ke dalam kluster ini. Dari sisi demografis nasabah, nasabah yang berasal dari provinsi DKI Jakarta memiliki propabilitas cukup besar untuk di klasifikasikan ke dalam kluster ini.

Diagram transisi keadaan dan matrik transisi keadaan *cluster 6* seperti yang terlihat pada gambar 27 dan tabel 20 menunjukkan sebagian besar fluktuasi saldo

rata-rata mingguan rekening-rekening yang tergabung ke dalam kluster ini berada pada level E. Dari matrik translasi tersebut juga bisa dilihat adanya fluktuasi statis dari saldo rata-rata mingguan dari rekening-rekening di kluster ini.

Jika diperhatikan lebih dalam dari matrik transisi tersebut terlihat adanya kecenderungan yang cukup tinggi (75%) nasabah yang memiliki saldo mingguan rata-rata pada level B, akan menutup rekening mereka dalam satu minggu ke depan.



Gambar 27. Diagram Transisi keadaan cluster 6

	A	B	C	D	E	X
A	0.33					0.67
B		0.25				0.75
C			0.40			0.40
D			0.03	0.79	0.03	0.15
E					0.96	0.03
X						0.83

Tabel 20. Matrik transisi keadaan cluster 6

7.2 Intisari pengetahuan yang didapat

Model binning5 yang digunakan untuk analisa lebih lanjut menyajikan beberapa informasi atau pengetahuan akan aspek-aspek yang mempengaruhi tutupnya sebuah rekening. Pengetahuan yang didapat tersebut perlu dibuat

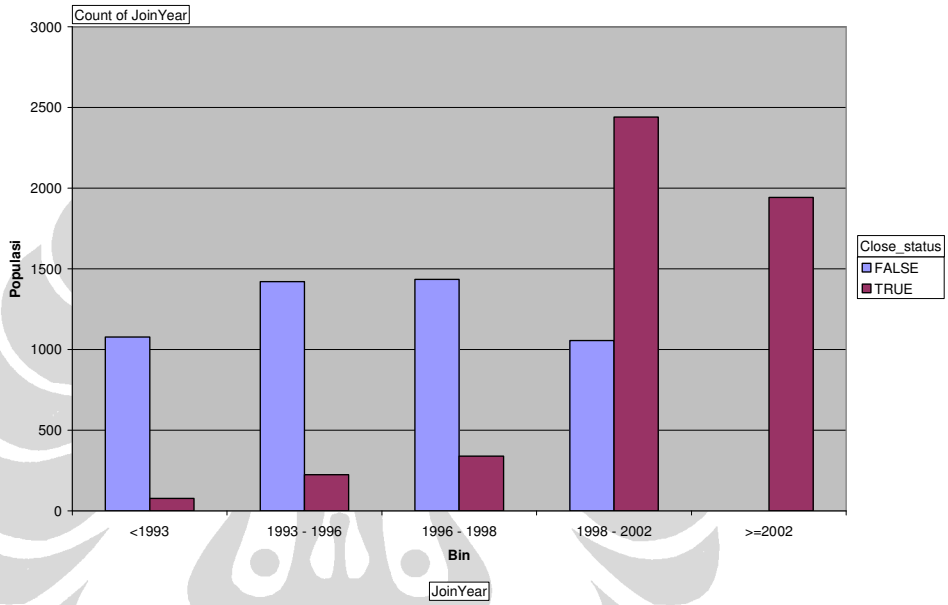
dirangkum lebih lanjut agar dapat dihadirkan sebagai pengetahuan yang bisa dipelajari kembali.

Berdasarkan analisa pada sub bab sebelumnya ditemukan bahwa pada kluster-kluster yang mengelompokkan rekening-rekening yang tutup (*cluster 1, cluster 3, cluster 4, cluster 5, cluster 6*), nasabah-nasabah pada kluster tersebut ditemukan enggan untuk memakai fasilitas kartu ATM. Hal sebaliknya terjadi pada *cluster 2* yaitu kluster yang merupakan kumpulan rekening-rekening aktif, nasabah-nasabah yang berada pada kluster ini memiliki kecenderungan lebih besar untuk memakai fasilitas kartu ATM.

Ketika penulis menelaah lebih dalam distribusi propabilitas dari variabel *Joinyear* seperti yang terlihat pada tabel 21 dibawah, penulis menemukan bahwa nasabah-nasabah yang bergabung tahun 1993 ke belakang (*Joinyear = < 1993*) cenderung bergabung ke dalam *cluster 2*. Sedangkan nasabah-nasabah yang bergabung lebih awal (*Joinyear = >= 2002*) kecil kemungkinannya untuk bergabung ke *cluster 2*. Hal ini berarti bahwa nasabah-nasabah yang bergabung tahun 2002 ke atas memiliki kecenderungan kuat untuk bergabung ke dalam kluster tempat berkumpulnya nasabah-nasabah yang rekeningnya sudah ditutup. Sedangkan nasabah-nasabah yang bergabung tahun 1993 ke belakang lebih cenderung untuk tidak menutup rekening mereka. Hal ini diperkuat juga dengan hasil analisa komposisi kelas variabel *close_status* untuk tiap interval, seperti yang terlihat pada gambar 28 dibawah

Joinyear	Cluster					
Class	1	2	3	4	5	6
>= 2002	49.34%	3.13%				
1998 - 2002	43.27%	24.93%	27.83%	20.00%	21.28%	26.09%
1996 - 1998	3.56%	34.55%	43.48%	46.67%	45.75%	40.58%
1993 - 1996	3.04%	22.49%	20.87%	26.67%	25.53%	23.19%
< 1993	0.79%	14.90%	7.83%	6.67%	7.45%	10.15%

Tabel 21. Distribusi propablitas variabel Joinyear

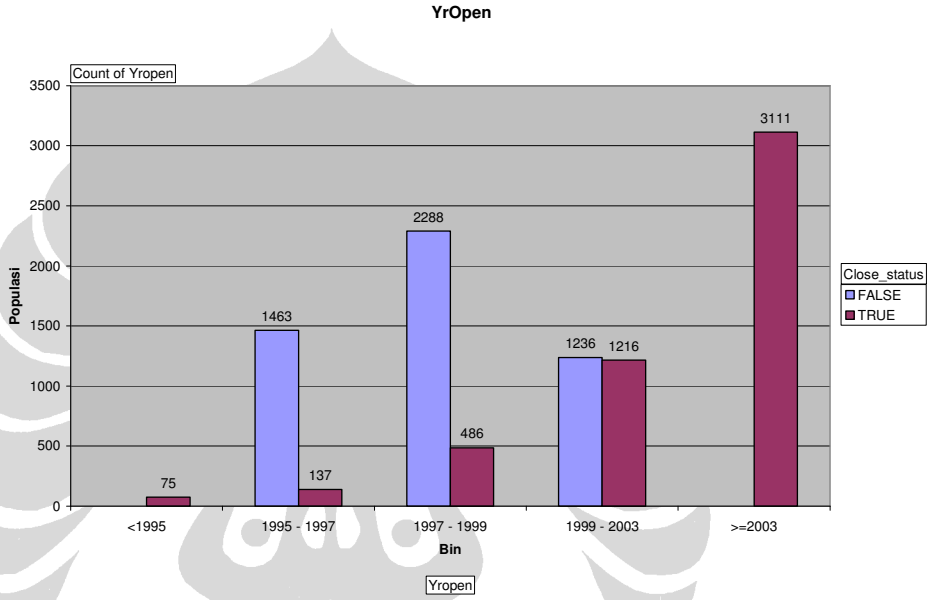


Gambar 28. Perbandingan komposisi Close_status untuk tiap bin variabel Joinyear

Penulis juga menelaah distribusi propabilitas dari variabel *Yropen* seperti yang terlihat pada tabel 22, berdasarkan tabel tersebut, penulis menemukan bahwa rekening-rekening yang dibuka tahun 2003 keatas (Yropen = >= 2003), kecil kemungkinannya untuk berada pada *cluster 2*. Hal ini ini memiliki arti bahwa rekening-rekening tersebut lebih cenderung bergabung ke dalam kluster-kluster yang memiliki kecenderungan sangat tinggi untuk tutup. Gambar 29 dibawah, yang menunjukkan komposisi kelas variabel *close_status* untuk tiap bin yang dipakai di variabel *Yropen*, memperkuat hal tersebut.

Yropen	Cluster					
Class	1	2	3	4	5	6
>= 2003	70.37%	6.44%	30.44%	13.33%	53.19%	21.74%
1999 - 2003	21.34%	22.42%	18.26%	26.67%	14.89%	21.74%
1997 - 1999	5.98%	40.62%	37.39%	33.33%	23.40%	43.48%
1995 - 1997	1.31%	30.38%	7.83%	26.67%	6.38%	10.15%
< 1995	1.00%		6.09%		2.13%	2.90%

Tabel 22. Distribusi propablitas variabel Yropen



Gambar 29. Perbandingan komposisi Close_status untuk tiap bin variabel Yropen

Melalui matrik transisi keadaan penulis mendapati bahwa kluster-kluster yang dengan status rekening tutup, memiliki fluktuasi saldo rata-rata mingguan yang statis. Statis disini berarti saldo mingguan rata-rata cenderung berfluktuasi pada level yang sama, sebagai contoh jika saat ini saldo mingguan berada di level "A", maka minggu depan rekening tersebut sangat besar kemungkinannya berada pada keadaan "A" juga. Kluster-kluster dengan status rekening masih aktif memiliki fluktuasi saldo rata-rata mingguan lebih dinamis. Dinamis disini berarti kecenderungan saldo rata-rata mingguan untuk berpindah ke level yang berbeda dengan level saat ini.