



UNIVERSITAS INDONESIA

Penggalian Pola Churn Menggunakan Data mining Pada
Institusi Perbankan Nasional

Disusun oleh

Thony Antonius

7205000733

Program Studi Magister Teknologi Informasi
Fakultas Ilmu Komputer
Universitas Indonesia
2008

LEMBAR PERSETUJUAN

Judul Tesis : PENGGALIAN POLA CHURN MENGGUNAKAN
DATA MINING PADA INSTITUSI PERBANKAN
NASIONAL

Nama : THONY ANTONIUS

NPM : 7205000733

TESIS INI TELAH DIPERIKSA DAN DISETUJUI

Jakarta, Juli 2008

Yudho Giri Sucahyo, Ph. D

PEMBIMBING

Kata Pengantar

Puji syukur ke hadirat Tuhan Yang Maha Kuasa, yang telah melimpahkan rahmat serta karunia-Nya sehingga Penulis dapat menyelesaikan Tesis yang berjudul Penggalan Pola Churn Menggunakan Data mining Pada Institusi Perbankan Nasional. Selanjutnya penulis ingin menyampaikan terima kasih kepada:

1. Bapak Yudho Giri S, Ph.D yang telah membimbing penulis selama mengerjakan tesis ini.
2. Seluruh dosen dan karyawan dari Magister Teknologi Informasi Universitas Indonesia, atas semua yang telah diberikan kepada penulis selama menjalani perkuliahan.
3. Teman-teman satu kelas angkatan 2005A atas dukungan moral dan semangat kebersamaan yang telah dibangun selama ini.
4. Kedua orang tuaku S. Tampubolon, dan E. Tambunan untuk semua cinta kasih yang telah diberikan.

Tesis ini dibuat untuk melengkapi persyaratan memperoleh gelar Magister pada Program Studi Magister Teknologi Informasi. Program Pascasarjana, Fakultas Ilmu Komputer, Universitas Indonesia. Semoga hasil dari tesis ini memiliki manfaat bagi semua pihak.

Abstrak

Kerasnya persaingan usaha dan suasana kompetisi pada industri perbankan yang semakin ketat menjadikan perusahaan berusaha sekeras mungkin untuk mencegah berpindahnya pelanggan mereka ke perusahaan pesaing. Salah satu cara untuk bisa mencegah berpindahnya pelanggan ke perusahaan pesaing adalah dengan melakukan prediksi dan deteksi dini pelanggan-pelanggan mana saja yang berpotensi meninggalkan perusahaan dan beralih ke perusahaan pesaing yaitu dengan melakukan *churn prediction*.

Churn prediction sudah diimplementasikan secara luas di industri telekomunikasi sebagai bagian dari *churn management*. Salah satu teknik yang digunakan untuk melakukan *churn prediction* adalah *data mining*. Tesis ini mencoba menggali pola-pola *churn* dari salah satu institusi perbankan nasional, dengan harapan bisa menemukan sebuah model *churn* bagi institusi perbankan tersebut. Hasil analisa yang dilakukan melahirkan pengetahuan mengenai kondisi seperti apa yang mengakibatkan seorang nasabah akan menutup rekening mereka. Penggalan informasi juga berhasil menemukan beberapa pola yang seperti apa yang bisa dijadikan pertanda seorang nasabah akan menutup rekening mereka.

Keterbatasan jumlah variabel dari *dataset* yang digunakan menghasilkan model *data mining* menjadi sangat sederhana, sehingga diperlukan adanya tambahan *variabel* lain untuk menghasilkan model yang lebih kuat.

Abstract

The harshness of the competition for efforts and the atmosphere of the competition in the increasingly tight banking industry made the company try as hard as possible to prevent their customer's moving to the competitor's company. Churn prediction is One of the methods that could prevent the customer's moving to the competitor's company by carrying out the prediction and the early detection of any customer who had the potential to leave the company and to change to the competitor's company.

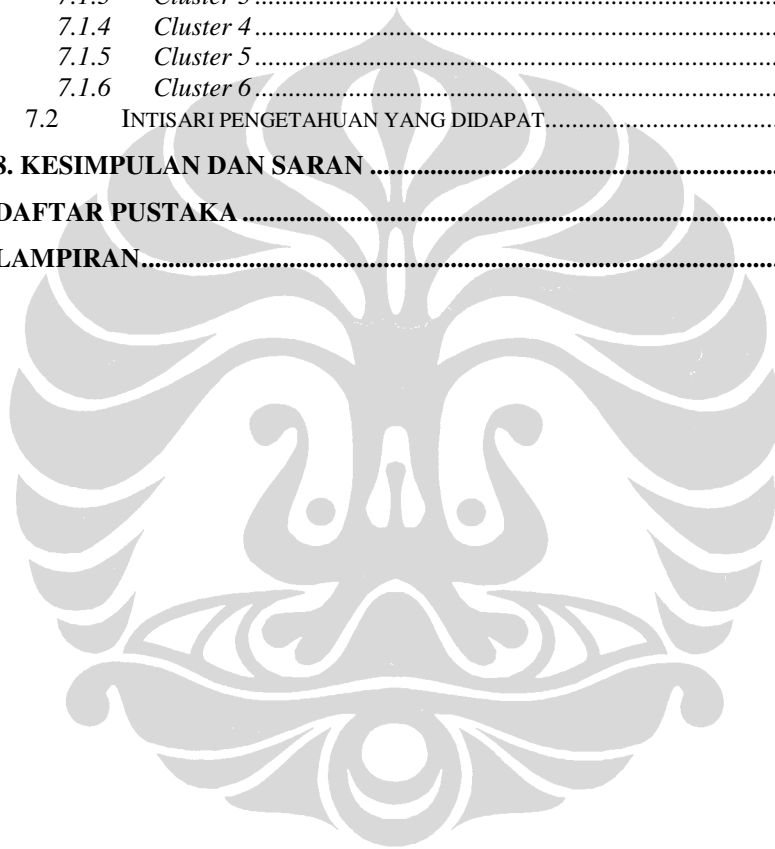
Churn prediction already implemented widely in the telecommunications industry as a part of churn management. One of the techniques that was used to do churn prediction was the data mining. This thesis tried to dig up patterns churn from one of the national banking institutions, in the hope of could find a model churn for this banking institution. Results of the analysis that was carried out produced knowledge concerning the condition like what resulted in a customer closing their account. The excavation of information also succeeded in finding several patterns that like what could be made the sign of a customer will close their account

The limitations of the number of variables from the set data that was used produced the data model mining became very simple, so as to be needed by the existence of the addition of the other variable to produce the stronger model

Daftar Isi

Kata Pengantar.....	i
Abstrak.....	ii
Abstract.....	iii
Daftar Isi.....	iv
Daftar gambar.....	vi
1. PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 PERMASALAHAN.....	5
1.3 TUJUAN DAN MANFAAT.....	5
1.4 LINGKUP PENELITIAN.....	6
1.5 STRUKTUR DOKUMEN.....	6
2. DASAR TEORI.....	8
2.1 DATA MINING.....	8
2.2 TEKNIK-TEKNIK DATA MINING.....	9
2.2.1 Clustering[5].....	10
2.2.2 Classification.....	20
2.3 EVALUASI MODEL.....	21
2.4 CHURN MANAGEMENT.....	23
3. METODOLOGI PENELITIAN.....	25
3.1 METODE DATA MINING.....	25
3.1.1 Fase business understanding.....	26
3.1.2 Fase data understanding.....	26
3.1.3 Fase data preparation.....	27
3.1.4 Fase modelling.....	27
3.1.5 Fase evaluasi.....	27
3.1.6 Fase deployment.....	27
3.2 OBYEK PENELITIAN.....	28
3.3 DATASET.....	28
3.4 SKEMA PELAKSANAAN PENELITIAN.....	29
4. PEMAHAMAN ISU BISNIS.....	30
4.1 GAMBARAN UMUM INDUSTRI PERBANKAN INDONESIA.....	30
4.2 PERSAINGAN KETAT DALAM INDUSTRI PERBANKAN.....	32
5. PEMAHAMAN DAN PERSIAPAN DATA.....	34
5.1 DATASET AWAL.....	35
5.2 EKSPLOKORASI DATASET AWAL.....	37
5.2.1 Variabel sex.....	37
5.2.2 Variabel Product.....	38
5.2.3 Variabel Custtype.....	39
5.2.4 Varibel CustClass.....	40
5.2.5 Varibel HasATMCard.....	41
5.2.6 Variabel ReferralBy.....	41
5.3 PERSIAPAN DATA.....	42
5.3.1 Atribut Province dan Dati II.....	44
5.3.2 Atribut dtopen, mtopen, yropen, week_open.....	44
5.3.3 Atribut tenuredays, tenureweek, tenuremonth.....	45
5.3.4 Tabel balance.....	45
5.3.5 Diskritisasi.....	46

5.4	PEMBENTUKAN TABEL YANG AKAN DI PROSES	50
5.4.1	<i>Tabel Acc_info</i>	50
5.4.2	<i>Tabel Balance_weekly_avg</i>	51
6.	PEMODELAN DAN EVALUASI.....	52
6.1	PEMBENTUKAN MODEL.....	53
6.2	PEMILIHAN ATRIBUT	54
6.3	EVALUASI MODEL	57
6.3.1	<i>Matrik klasifikasi</i>	57
7.	EMAS YANG DIPEROLEH.....	61
7.1	ANALISA LEBIH LANJUT MODEL BINNING5	61
7.1.1	<i>Cluster 1</i>	63
7.1.2	<i>Cluster 2</i>	64
7.1.3	<i>Cluster 3</i>	66
7.1.4	<i>Cluster 4</i>	67
7.1.5	<i>Cluster 5</i>	69
7.1.6	<i>Cluster 6</i>	70
7.2	INTISARI PENGETAHUAN YANG DIDAPAT.....	71
8.	KESIMPULAN DAN SARAN	75
	DAFTAR PUSTAKA	78
	LAMPIRAN.....	80



Daftar gambar

Gambar 1. Masalah yang dihadapi dalam penggunaan clustering[5]	11
Gambar 2. Perbedaan metode single-link dengan metode Complete-link[5]	14
Gambar 3. Distribusi Propabilitas perpindahan antara keadaan satu ke keadaan lain.....	18
Gambar 4. Rantai Markov [10]	19
Gambar 5. Matrix transisi keadaan	20
Gambar 6. Model data mining CRISP-DM	25
Gambar 7. Arsitektur Datawarehouse	28
Gambar 8. Tahapan pelaksanaan data mining.....	29
Gambar 9. Distribusi data variabel Sex.....	38
Gambar 10. Distribusi data variabel Product	39
Gambar 11. Distribusi variabel custtype.....	40
Gambar 12. Distribusi Variabel CustClass	40
Gambar 13. Rasio Kepemilikan Kartu ATM	41
Gambar 14. Sebaran data variabel ReveralBy	42
Gambar 15. Sampling acak field balance dari tabel balance_week_avg	47
Gambar 16. Sebaran data diskritisasi jenis 5	48
Gambar 17. Sebaran data diskritisasi jenis 7	49
Gambar 18. Sebaran data diskritisasi jenis 9	50
Gambar 19. Grafik distribusi pembagian dataset train dan test	53
Gambar 20. Pemilihan Variabel.....	55
Gambar 21. Diagram kluster model binning5.....	62
Gambar 22. Diagram transisi keadaan cluster 1.....	64
Gambar 23. Diagram transisi keadaan cluster 2.....	65
Gambar 24. Diagram Transisi keadaan cluster 3	67
Gambar 25. Diagram Transisi keadaan cluster 4	68
Gambar 26. Diagram Transisi keadaan cluster 5	70
Gambar 27. Diagram Transisi keadaan cluster 6	71
Gambar 28. Perbandingan komposisi Close_status untuk tiap bin variabel Joinyear	73
Gambar 30. Perbandingan komposisi Close_status untuk tiap bin variabel Yropen	74