

# PERANCANGAN SISTEM PREDIKSI *CHURN* PELANGGAN PT. TELEKOMUNIKASI SELULER DENGAN MEMANFAATKAN PROSES *DATA MINING*

Rajesri Govindaraju<sup>1</sup>, Tota Simatupang<sup>2</sup>, TMA. Ari Samadhi<sup>3</sup>

<sup>1,2</sup>Kelompok Keahlian Sistem Industri & Tekno-Ekonomi, Fakultas Teknologi Industri, Institut Teknologi Bandung

<sup>3</sup>Kelompok Keahlian Sistem Manufaktur, Fakultas Teknologi Industri, Institut Teknologi Bandung

E-mail: rajesri\_g@mail.ti.itb.ac.id<sup>1</sup>

**ABSTRAK:** Penelitian ini bertujuan merancang sistem prediksi *churn* pelanggan yang memanfaatkan proses data mining. Sistem yang dihasilkan dapat melakukan integrasi data, pembersihan data, transformasi data, sampling dan pemisahan data, konstruksi model prediksi, memprediksi *churn* pelanggan dan menampilkan hasil prediksi dalam format laporan tertentu yang diperlukan. Identifikasi variabel-variabel prediksi *churn* dilakukan berdasarkan model prediksi *churn* yang telah dikembangkan pada penelitian terdahulu yang antara lain mencakup informasi mengenai pelanggan, metode pembayaran, data percakapan, data penggunaan jenis-jenis layanan telekomunikasi dan data yang menggambarkan perubahan perilaku penggunaan layanan telekomunikasi tersebut. Teknik *mining* yang dipilih adalah teknik klasifikasi dengan algoritma *decision tree*. *Decision tree* menghasilkan model visual yang merepresentasikan pola perilaku pelanggan yang *churn* dan tidak *churn*. Uji coba sistem yang dilakukan menggunakan data pelanggan Kartu Halo daerah Bandung menghasilkan tingkat akurasi model prediksi sebesar 70,94%.

**Kata kunci:** *customer relationship management (CRM)*, *churn*, *data mining*, *decision tree*, sistem prediksi *churn*.

**ABSTRACT:** The purpose of this research is to design a customer churn prediction system using data mining approach. This system is able to perform data integration, data cleaning, data transformation, sampling and data splitting, prediction model building, predicting customer churn, and show the results in certain agreed forms. Churn prediction variables were identified based on earlier research reports that include customer information, payment method, call pattern, complaint data, telecommunication services usage and change of telecommunication services usage behavior data. The preferred mining technique used is the classification with decision tree algorithm. The decision tree can present visual model which represents customer churn and non churn pattern behavior. This system was tested using Kartu Halo customer data in Bandung area and testing result showed 70,94% accuracy of the prediction model.

**Keywords:** *customer relationship management (CRM)*, *churn*, *data mining*, *decision tree*, *churn prediction system*.

## PENDAHULUAN

Salah satu tantangan yang dihadapi CRM perusahaan operator telekomunikasi seluler adalah usaha menurunkan jumlah pelanggan yang berhenti menggunakan layanan perusahaan dan pindah ke perusahaan kompetitor. Perilaku pelanggan yang meninggalkan layanan yang diberikan perusahaan pada industri operator telekomunikasi seluler disebut *churn*. Rata-rata *churn rate* per bulan untuk perusahaan telekomunikasi seluler di Eropa berkisar antara 8% sampai 12% dan biaya *churn*-nya sekitar 500 euro [1]. Sementara di Amerika Serikat, *churn rate* domestik per bulan pada tahun 1998 adalah 2-3 %. Biaya rata-rata untuk memperoleh seorang pelanggan baru mencapai 400 dolar dan biaya *churn* pada industri ini mendekati 6,3 juta dolar. Total kerugian per tahun dapat meningkat menjadi 9,6 juta dolar apabila

kerugian pendapatan per bulan dari pelanggan juga ikut diperhitungkan, sehingga pada suatu perusahaan yang memiliki 1,5 juta pelanggan, pengurangan *churn rate* dari 2% menjadi 1% dapat meningkatkan pendapatan tahunan sedikitnya sebanyak 54 juta dolar dan meningkatkan *shareholder value* sampai 150 juta dolar [2]. Berdasarkan fakta tersebut, *churn* merupakan proses yang dapat mengurangi keuntungan perusahaan. Oleh karena itu, manajemen *churn* menjadi senjata yang krusial dalam berkompetisi, dan suatu fondasi pada suatu usaha pemasaran berorientasi pelanggan.

Sebagai perusahaan telekomunikasi yang menyediakan jasa operator telepon seluler, PT Telekomunikasi Seluler (Telkomsel) telah menyadari pentingnya mengatasi masalah *churn*. Perusahaan ini merupakan pemimpin pasar industri telepon seluler dengan *market share* lebih dari 50%. Dengan *market*

*share* yang besar tersebut, tentunya perlu dilakukan usaha-usaha agar setidaknya nilai *market share* yang telah diraih tidak menurun, sehingga perlu dilakukan cara yang efisien untuk mempertahankan *market share*, dalam artian cara yang akan mengeluarkan sedikit biaya, namun hasil yang signifikan. Karena ongkos mempertahankan pelanggan 5 kali lebih murah daripada memperoleh pelanggan baru [3] maka mempertahankan pelanggan dan mengurangi *churn rate* merupakan cara yang paling tepat bagi PT Telekomunikasi Seluler untuk mempertahankan *market share*-nya.

Berdasarkan data laporan 31 Januari 2005 untuk regional Jawa Barat dan Produk Kartu Halo pada tabel diatas, jumlah pelanggan yang *churn* sebesar 3160, sedangkan pelanggan yang masih aktif sebesar 111252. Besarnya jumlah pelanggan yang *churn* tersebut mengindikasikan bahwa masih terbuka peluang yang besar untuk mempertahankan *market share* dengan cara mencegah pelanggan-pelanggan yang aktif menjadi *churn*. Agar dapat melakukan tindakan-tindakan preventif untuk mencegah *churn*, Divisi CRM perusahaan perlu mengetahui dan memahami pola perilaku *churn* pelanggannya, serta memprediksi apakah pelanggan-pelanggan aktif yang dimiliki saat ini akan *churn* di masa depan. Hasil prediksi ini akan menjadi informasi yang berguna dalam merancang dan memutuskan program retensi yang akan dilakukan. Dengan memprediksi pelanggan yang akan *churn*, Telkomsel dapat mengurangi *churn rate* dengan menawarkan program-program retensi baru yang membuat pelanggan tetap bertahan. Dengan memahami mengapa pelanggan dapat *churn*, Telkomsel juga dapat berusaha untuk memperbaiki layanan-layanannya untuk memuaskan pelanggan di masa depan.

Telkomsel yang mempunyai sekitar 24 juta pelanggan, telah memiliki basis data untuk menyimpan data-data yang dibutuhkan dalam menjalankan bisnisnya. Data-data berjumlah besar yang tersimpan secara elektronik pada basis data Telkomsel dapat dimanfaatkan untuk menemukan pola-pola perilaku dan karakteristik konsumen. Pengolahan yang tepat pada data-data tersebut dapat menghasilkan pengetahuan-pengetahuan yang bermanfaat untuk memahami pola perilaku *churn* serta memprediksi pelanggan mana yang akan meninggalkan perusahaan. Proses analisis data untuk menemukan informasi dan pengetahuan pada data yang sangat besar disebut *data mining*. *Data mining* sebagai salah satu alat analisis pada *analytical CRM* telah terbukti kegunaannya bagi suatu perusahaan dalam memprediksi pelanggan yang akan meninggalkan perusahaan, kapan mereka akan *churn*, dan tingkat akurasi prediksi tersebut [4]. Oleh

karena itu, *data mining* dapat digunakan sebagai pendekatan untuk menjawab kebutuhan analisis dan prediksi *churn* dalam memecahkan masalah *churn* pada Telkomsel.

Berdasarkan uraian tersebut serta mengingat bahwa prediksi *churn* merupakan aktivitas yang dilakukan berulang-ulang, dapat ditarik kesimpulan bahwa Telkomsel membutuhkan suatu sistem prediksi *churn* untuk menyediakan pengetahuan dan informasi yang mendukung pembuatan program-program retensi serta mengurangi jumlah pelanggan yang *churn*. Proses *data mining* dapat digunakan sebagai alat analisis dan prediksi *churn* pelanggan yang merupakan salah satu fitur *analytical CRM*. Dengan mempertimbangkan besarnya jumlah data yang digunakan, analisis dan prediksi *churn* akan menghabiskan banyak waktu dalam pelaksanaannya. Ditambah dengan kebutuhan akan akurasi dan kecepatan pemrosesan data, serta kelancaran aliran informasi maka perlu dikembangkan suatu aplikasi perangkat lunak yang dapat melaksanakan fungsi-fungsi sistem prediksi *churn* yang dirancang.

## LANDASAN TEORI

Pada bagian ini akan diulas secara singkat teori dan metode utama yang digunakan dalam penelitian ini, yaitu teori mengenai variabel prediksi *churn* dan metode Decision Tree.

### Variabel Prediksi Customer Churn

Baragoin, C. et al [5] menyatakan bahwa sumber data yang dapat digunakan untuk memprediksi pelanggan yang *churn* pada perusahaan operator layanan telekomunikasi seluler adalah sebagai berikut:

- Indikator *Churn*. Dalam memprediksi *churn*, dibutuhkan variabel yang akan diprediksi. Pada kasus ini variabel ini berisi data apakah pelanggan *churn* atau tidak.
- Data Informasi Pelanggan
  - Data Demografi. Berisikan data-data yang dibawa oleh pelanggan saat pertama kali mereka memulai menggunakan layanan telepon dan mendeskripsikan karakteristik mereka.
  - Data Kontrak. Berisikan data-data kontrak pelanggan yang diperoleh saat kontrak dimulai.
- *Data Call - Data call* dapat dideskripsikan sebagai perilaku berkomunikasi (*incoming* atau *outgoing*) pelanggan dalam berbagai cara.
- *Billing* dan Data Pembayaran
- Indeks Pelanggan yang Diturunkan dari Data Transaksi
- Data Tambahan

Penelitian lain oleh [6] mencoba memprediksi *churn* dengan menggunakan data *complaints* dan perbaikan untuk memprediksi *customer churn*. Tiga grup variabel dikombinasikan untuk menghasilkan set data sebagai berikut:

- *Provisions Data*, data ini merepresentasikan estimasi-estimasi yang dibuat oleh perusahaan mengenai resolusi dari *complaint* atau perbaikan.
- *Complaints Data*, data ini merepresentasikan informasi mengenai *complaint*.
- *Repairs Data*, data ini merepresentasikan semua informasi mengenai kesalahan dan perbaikan.

Sementara [7] menggunakan sejumlah variabel untuk menganalisis dan memprediksi *churn* yang antara lain mencakup pendapatan, caller, gender, age, handset, peyment method, lama percakapan, dan lain-lain.

**Decision Tree**

*Decision tree* adalah suatu *flowchart* seperti struktur pohon, dimana tiap titik internalnya (*internal node*) menunjukkan suatu *test* pada suatu atribut, tiap cabang (*branch*) merepresentasikan hasil dari *test* tersebut, dan *leaf node* menunjukkan kelas-kelas atau distribusi kelas [8]. *Node* yang paling diatas adalah *root node*. Suatu *decision tree* yang terlihat seperti pada Gambar 1 merupakan *decision tree* yang merepresentasikan konsep membeli komputer, yaitu untuk memprediksi apakah seorang pelanggan di AllElectronics akan membeli suatu komputer atau tidak. *Node* internal ditunjukkan dengan segi empat, dan *leaf node* ditunjukkan dengan oval.

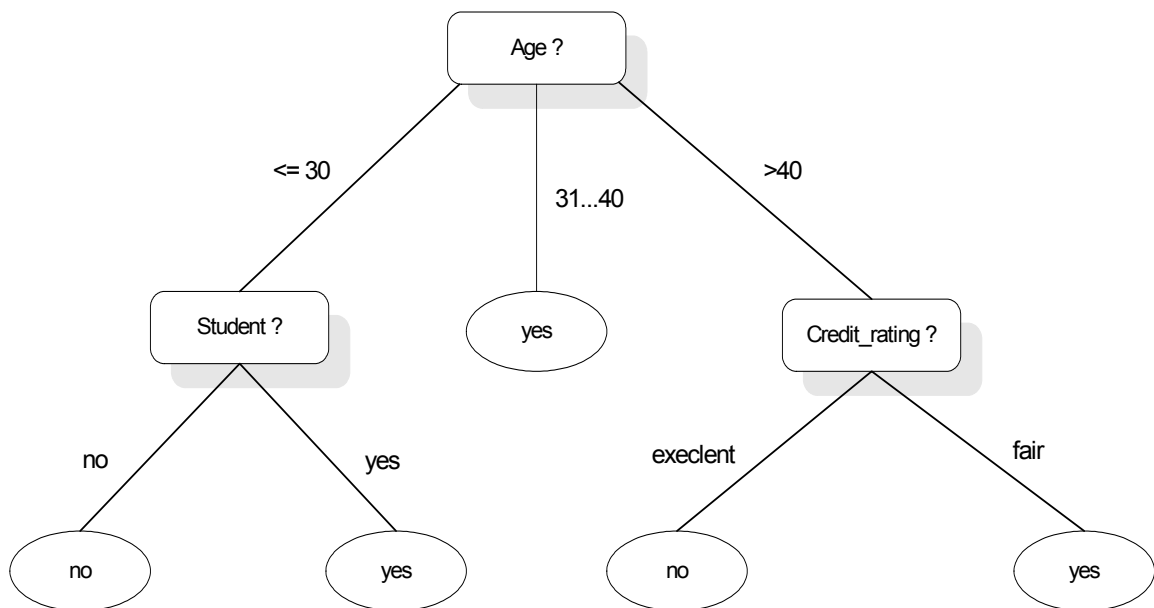
Untuk mengklasifikasikan sampel yang tidak diketahui, nilai atribut dari sampel tersebut dites oleh *decision tree*. Suatu jalur ditelusuri dari *root* ke *node leaf* yang memiliki prediksi kelas untuk sampel tersebut. *Decision tree* dengan mudah dapat diubah dari aturan (*rule*) prediksi ke klasifikasi. Saat *decision tree* dibuat, kebanyakan dari cabang mungkin memperlihatkan *noise* atau *outlier* pada *training data*. *Tree pruning* berusaha untuk mengidentifikasi dan membuang cabang-cabang tersebut dengan tujuan memperbaiki akurasi klasifikasi pada data yang tidak kelihatan. *Decision tree* telah digunakan dalam berbagai area aplikasi dari kedokteran sampai *game theory* dan bisnis.

Berikut ini adalah algoritma dasar untuk menghasilkan *decision tree* dari *sample training* [8]:

- **Algorithm:** *generate\_decision\_tree*. Menghasilkan *decision tree* dari *training data* yang diberikan.
- **Input:** sampel *training*, sampel, direpresentasikan dengan atribut bernilai diskrit, atribut kandidat, *attribut-list*.
- **Output:** suatu *decision tree*

```

Method:
(1) create a node N;
(2) if samples are all of the same class, C then
(3) return N as a leaf node labeled with class C;
(4) if attribute-list is empty then
(5) return N as a leaf node labeled with the most
    common class in samples; // majority voting
(6) select test-attribute, the attribute among attribute-
    list with the highest information gain;
(7) label node N with test attribute;
    
```



Gambar 1. *Decision Tree* Pembelian Komputer (Han dan Kamber, 2001)

(8) for each known value  $a_i$  of test-attribute // partition the samples  
 (9) grow a branch from node  $N$  for the condition test-attribute =  $a_i$ ;  
 (10) let  $si$  be the set of samples for which test-attribute =  $a_i$ ; // a partition  
 (11) if  $si$  is empty then  
 (12) attach a leaf labeled with the most common class in samples; else attach the node returned by `Generate_decision_tree( $si$ , attribute-list-test-attribute)`;

Algoritma dasar untuk induksi *decision tree* adalah satu algoritma yang mengkonstruksi *decision tree* dalam cara pembagian *top-down recursive*. Algoritma yang telah dibahas sebelumnya adalah algoritma versi ID3 yang merupakan algoritma induksi *decision tree* yang terkenal. Strategi dasar dari algoritma tersebut adalah sebagai berikut:

- *Tree* dimulai dengan suatu *node* merepresentasikan sampel *training* (langkah 1)
- Jika semua sampel berada dalam satu kelas, maka *node* menjadi sebuah *leaf* dan diberi label berdasarkan kelas tersebut (langkah 2 dan 3)
- Jika tidak, algoritma akan menggunakan pengukuran berdasarkan entropi yang dikenal dengan *information gain* sebagai suatu pendekatan heuristik untuk menyeleksi atribut yang terbaik dalam memisahkan sampel kedalam kelas individual (langkah 6). Atribut ini menjadi "*test*" atau "*decision*" atribut pada *node* (langkah 7). Dalam versi algoritma ini, semua atribut adalah kategorial, yaitu bernilai diskrit. Atribut nilai kontinyu harus dideskripsikan.
- Suatu *branch* dibuat untuk setiap nilai yang diketahui pada atribut yang diuji, dan sampel dipartisi berdasarkan nilai atribut tersebut.
- Algoritma menggunakan proses yang sama secara berulang-ulang untuk membentuk *decision tree* untuk tiap sampel pada tiap partisi. Sewaktu atribut sudah berada di *node*, tidak diperlu diperhitungkan lagi turunannya (langkah 13).
- Partisi yang berulang-ulang berhenti hanya jika salah satu kondisi bernilai benar:
  - Semua sampel pada *node* yang diberikan memiliki kelas yang sama (langkah 2 dan 3), atau
  - Tidak ada atribut yang tersisa dimana sampel dipartisi lebih jauh lagi (langkah 4). Dalam kasus ini dilakukan voting mayoritas (langkah 5). Hal ini melibatkan perubahan *node* yang ada ke *leaf* serta menamainya dengan kelas

mayoritas diantara sampel. Alternatif yang lain adalah tetap menyimpan distribusi kelas dari *node* sampel.

- Tidak ada sampel untuk *branch test-attribute* =  $a_i$  (langkah 11). Dalam kasus ini, sebuah *leaf* dibuat dengan kelas mayoritas dalam sampel.

## ANALISIS KEBUTUHAN

Kebutuhan bisnis dikategorikan menjadi kebutuhan fungsional dan kebutuhan non-fungsional [9]. Analisis kebutuhan sistem dilakukan berdasarkan literatur dan wawancara terhadap pihak Telkomsel. Kebutuhan fungsional merupakan deskripsi dari aktivitas-aktivitas dan layanan-layanan yang dapat diberikan oleh sistem, yang penelitian ini adalah sistem prediksi *churn*. kebutuhan fungsional sistem prediksi *churn* dalam penelitian ini adalah:

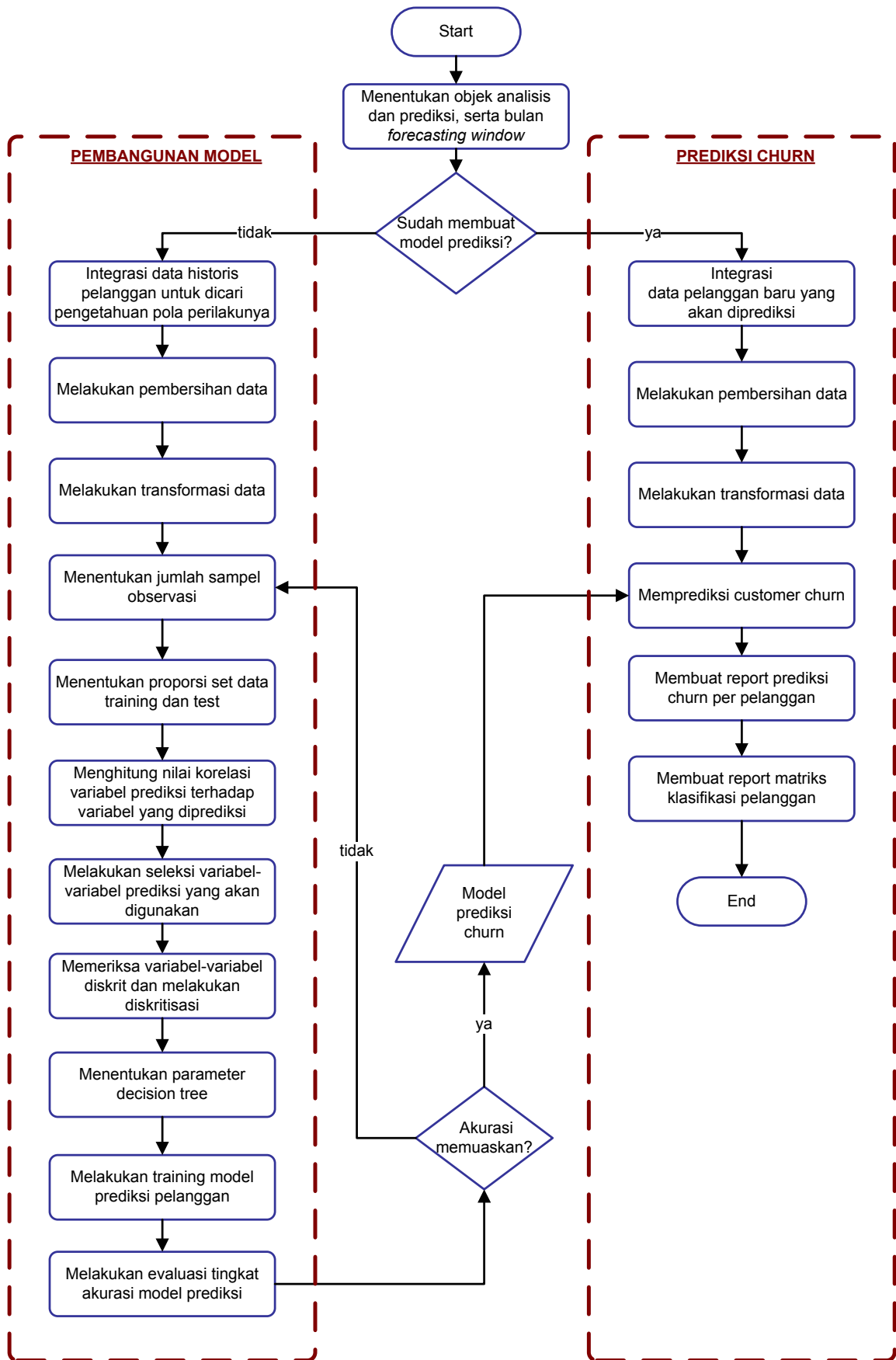
- Dapat mengintegrasikan data-data yang berasal dari sumber data Telkomsel.
- Dapat menyimpan data yang dibutuhkan dalam bentuk tabel atau *view* yang terstruktur.
- Dapat menyiapkan data agar siap digunakan untuk *input* algoritma *data mining*.
- Dapat menjalankan fungsi-fungsi utama *data mining*, yaitu membangun model prediksi dan melakukan prediksi *churn*.
- Dapat menampilkan laporan hasil prediksi *churn* per pelanggan dan matriks klasifikasi pelanggan.

Kebutuhan non-fungsional merupakan deskripsi dari fitur-fitur lain, karakteristik, dan *constraint* yang didefinisikan untuk memenuhi kebutuhan sistem. Kebutuhan non-fungsional sistem adalah:

- Dapat menyimpan dan mengolah data berukuran besar.
- Dapat menghasilkan *output* yang diinginkan secara cepat dan akurat.

## PERANCANGAN SISTEM

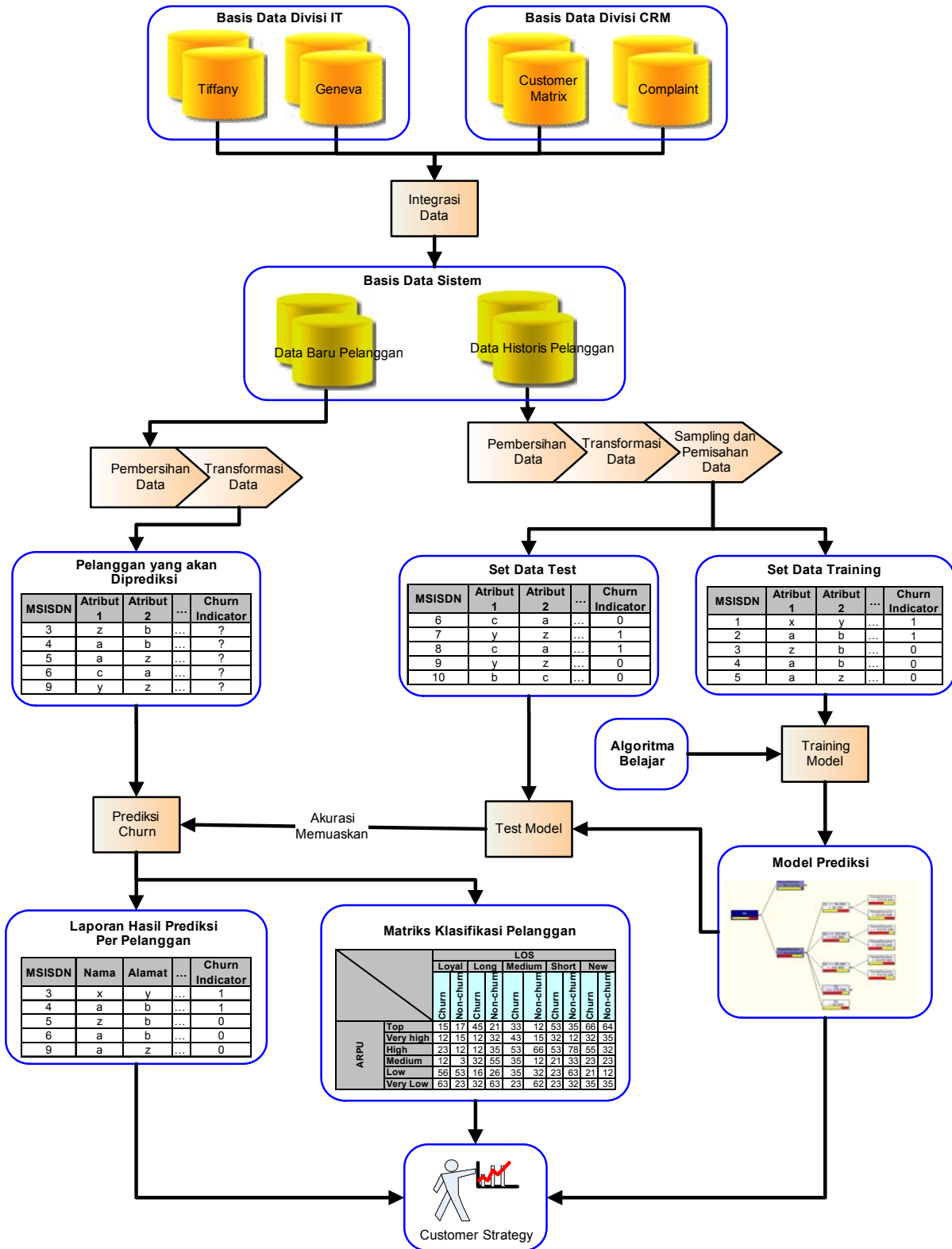
Sebelum merancang sistem prediksi *churn*, perlu dilakukan pemahaman proses-proses analisis dan prediksi *churn*. Selanjutnya, proses-proses tersebut disusun dan direpresentasikan dalam bentuk *flowchart* seperti pada Gambar 2. Terdapat dua proses besar pada *flowchart*, yaitu tahap pembangunan model dan tahap prediksi *churn*. Tahap pembangunan model akan menghasilkan *output* berupa model prediksi *churn*. Selanjutnya, model prediksi *churn* yang dihasilkan akan digunakan pada tahap prediksi *churn* sehingga menghasilkan *output* dalam bentuk laporan hasil prediksi *churn* per pelanggan dan matriks klasifikasi pelanggan.



Gambar 2. Proses Analisis dan Prediksi Churn

Selanjutnya dirancang sistem prediksi *churn* yang dapat menjalankan proses analisis dan prediksi *churn*, serta kebutuhan fungsional dan kebutuhan non-fungsional yang telah didefinisikan. Sistem prediksi

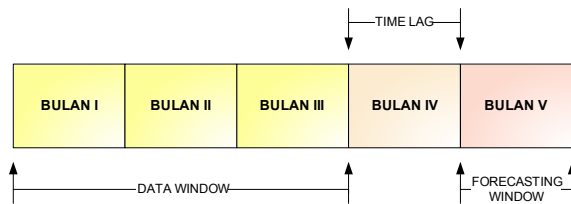
*churn* yang dirancang dapat dilihat pada Gambar 3. Kemudian akan dijelaskan secara lebih rinci bagian-bagian dari sistem tersebut.



Gambar 3. Sistem Prediksi Churn

**Penentuan Kerangka Waktu Prediksi Churn**

Komponen-komponen kerangka waktu yang digunakan untuk analisis dan prediksi *churn* terdiri dari *data window*, *time lag*, dan *forecasting window* seperti yang diperlihatkan Gambar 4. *Data Window* merupakan kerangka waktu untuk variabel *input* yang digunakan untuk mengkonstruksi model. Penelitian ini menggunakan *data window* selama 3 bulan dengan mempertimbangkan usaha yang cukup besar dan waktu yang lama dalam *sourcing* dan *preprocessing* data apabila *data window* semakin panjang.



**Gambar 4. Kerangka Waktu Prediksi Churn**

*Time lag* merupakan waktu yang digunakan untuk melakukan prediksi serta melakukan perencanaan dan eksekusi program *marketing*. *Time lag* yang digunakan adalah 1 bulan, karena berdasarkan hasil wawancara waktu selama 1 bulan cukup untuk digunakan dalam melakukan prediksi *churn*, merencanakan program retensi dan mengeksekusi program retensi. *Forecasting window* menggambarkan rentang waktu perilaku *churn* pelanggan yang akan diprediksi. *Churn* atau tidaknya pelanggan ditentukan dari *churn* atau tidaknya pelanggan pada *forecasting window*. *Forecasting window* yang digunakan dalam penelitian ini adalah satu bulan, sehingga dalam satu tahun dilakukan 12 kali prediksi *churn*.

**Model Data Prediksi Churn**

Model data [5] digunakan sebagai kerangka dasar dari model data prediksi *churn* yang digunakan dalam penelitian ini. Model data [5] ini dilengkapi dengan menambahkan variabel-variabel prediksi *churn* yang berasal dari penelitian [6] serta penelitian [7]. Model data prediksi *churn* yang digunakan terdiri dari 61 atribut, yang meliputi:

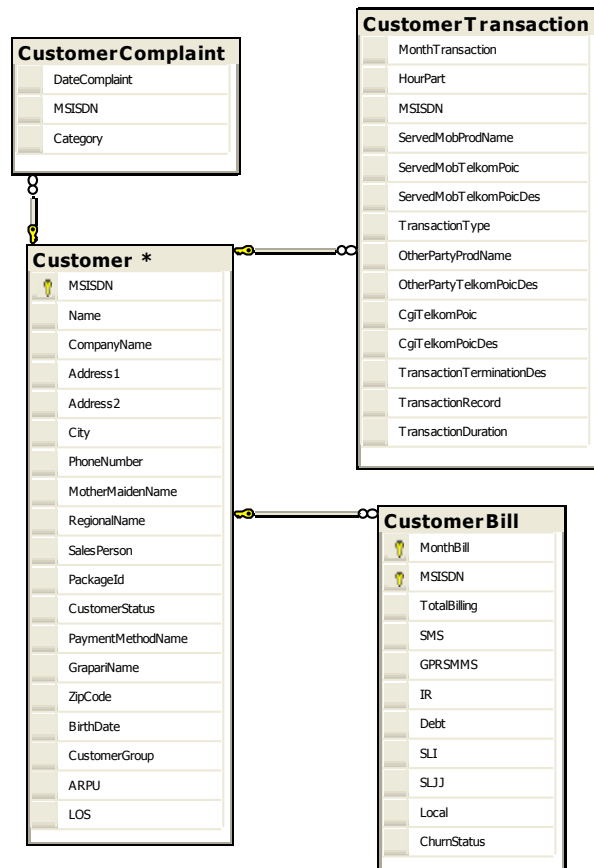
- *Key Atributte*, yaitu MSISDN.
- *Predictable Atributte*, yaitu ChurnIndicator.
- Atribut yang merupakan variabel prediksi *churn*, yaitu:
  - Data pelanggan, misalnya Age, LOS, CustomerGroup, dan lain-lain.
  - *Billing*, misanya ARPU, SMS, SLI, IR, Local, dan lain-lain.
  - *Data Complaint*.
  - *Call data*, misalnya Call Duration, Peak Call Duration, CompleteCall, SMS Record, Out-sphere Call dan lain-lain.

- Indeks turunan, misalnya Mobility, SMS Trend, Quality Call, dan lain-lain.

**Basis Data Sistem**

Data-data yang digunakan disimpan dan distrukturkan dalam bentuk basis data relasional. Basis data yang terstruktur ini dapat memudahkan manajemen data, seperti mengambil data untuk diolah dan menyimpan hasil olahannya. Struktur basis data itu digambarkan dalam bentuk diagram *conceptual data model* pada Gambar 5. Tabel-tabel yang terdapat pada *conceptual data model* berisi data-data yang diambil dari berbagai sumber data. Isi tabel-tabel tersebut dijelaskan sebagai berikut:

- Tabel Customer menyimpan data dan informasi mengenai pelanggan yang diperkirakan tidak banyak berubah dari waktu ke waktu.
- Tabel CustomerTransaction menyimpan transaksi yang dilakukan oleh pelanggan selama 3 bulan *data window*.
- Tabel CustomerBill menyimpan data billing pelanggan yang terjadi selama 3 bulan *data window* dan status *churn* masing-masing pelanggan pada *forecasting window*.
- Tabel CustomerComplaint menyimpan data *complaint* yang pernah dilakukan pelanggan selama 3 bulan *data window*.



**Gambar 5. Basis Data Sistem**



**Integrasi data**

Sumber data yang digunakan pada penelitian ini hanyalah sumber data internal, sehingga beberapa variabel yang berdasarkan model penelitian tertentu tidak bisa diperoleh untuk memprediksi *customer churn*. Data-data yang berasal dari berbagai sumber internal, yaitu divisi dan aplikasi pada Telkomsel diintegrasikan ke dalam suatu basis data yang terdiri dari tabel-tabel yang saling berhubungan seperti yang diperlihatkan pada Gambar 6.

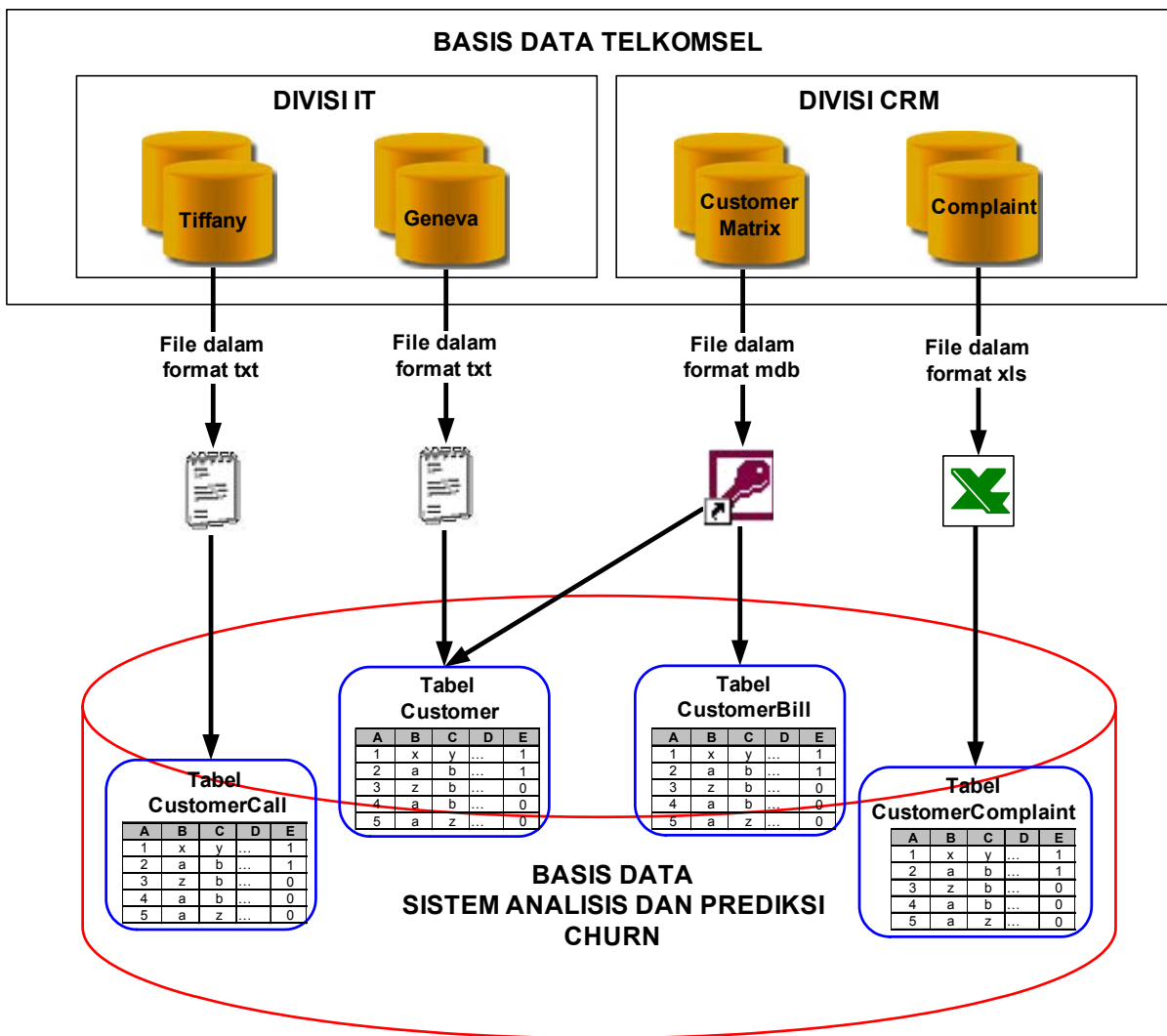
**Pemilihan Teknik Data Mining**

Pada penelitian ini algoritma klasifikasi dan prediksi yang dipilih adalah *decision tree*. Alasan penggunaan *decision tree* adalah:

- *Decision tree* merupakan metode representasi logika keputusan yang umum digunakan dalam memprediksi *churn*. Hal ini dibuktikan dengan

banyaknya penelitian prediksi *churn* pelanggan yang menggunakan *decision tree* sebagai model prediksi yang kemudian menghasilkan tingkat akurasi yang memuaskan, misalnya penelitian [6] dan penelitian [7]. Penelitian-penelitian tersebut membuktikan bahwa *decision tree* merupakan metode yang cukup baik dan stabil dalam memprediksi *churn*. *Decision tree* menghasilkan visualisasi pohon keputusan yang merupakan representasi logika-logika keputusan mengapa suatu pelanggan diprediksi akan *churn*.

- *Decision tree* menghasilkan pohon keputusan yang merupakan representasi logika-logika keputusan mengapa suatu pelanggan diprediksi akan *churn* dan mengapa suatu pelanggan diprediksi tidak akan *churn*. Hal ini sangat berguna dalam memahami dan menganalisis pola perilaku *churn* pelanggan.



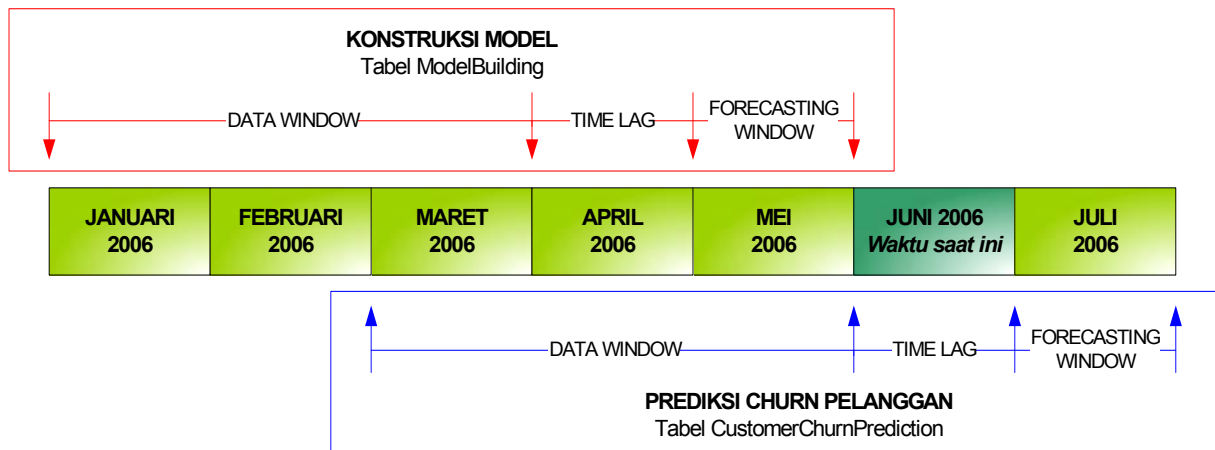
Gambar 6. Diagram Mekanisme Sourcing



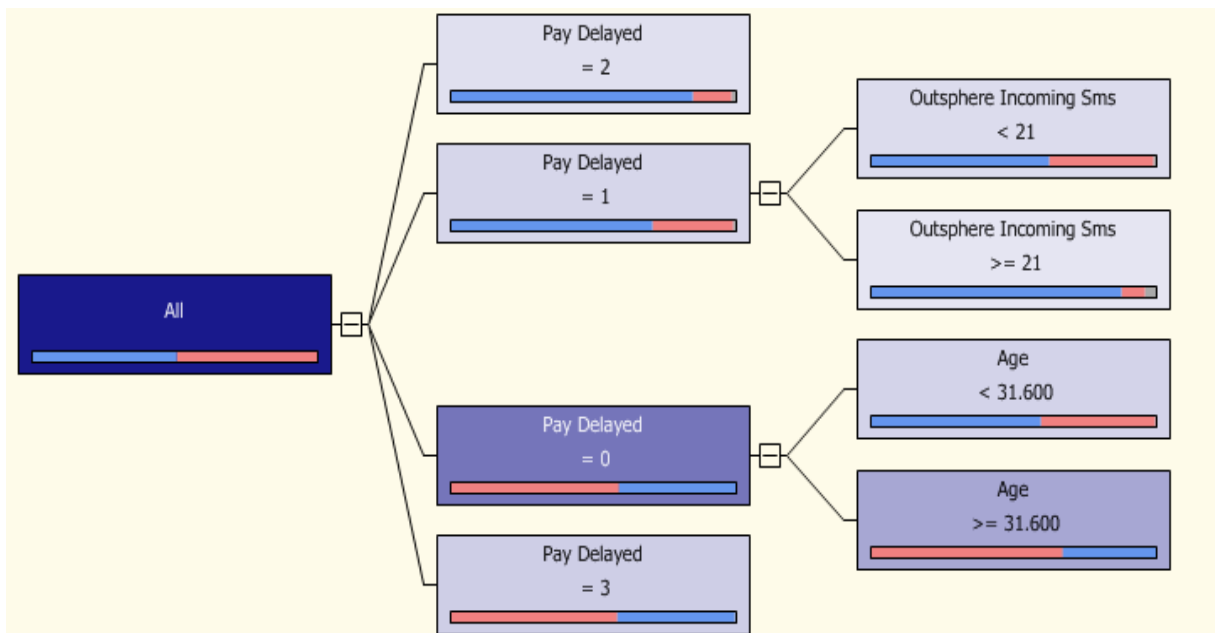
**UJICоба SISTEM**

Ujicoba sistem dilakukan untuk menguji apakah sistem prediksi *churn* ini dapat dijalankan sesuai dengan yang diharapkan. Kasus yang diambil adalah analisis pola perilaku pelanggan dan prediksi *churn* pelanggan Kartu Halo yang berada di kota Bandung. Data-data digunakan mengikuti kerangka waktu yang telah ditentukan pada tahap perancangan. Untuk data *historis* pelanggan yang akan digunakan untuk membangun model prediksi, digunakan *data window* yang berawal dari bulan Januari 2006 sampai bulan Maret 2006, *time lag* pada April 2006, serta *forecasting window* pada bulan Mei 2006 seperti yang dapat dilihat pada Gambar 7.

*Training* model menghasilkan model prediksi *decision tree* yang dapat dilihat pada Gambar 8. Selanjutnya model prediksi tersebut digunakan untuk memprediksi apakah pelanggan yang dimiliki Telkomsel saat ini akan *churn* atau tetap aktif pada bulan Juli 2006. Tingkat akurasi yang dihasilkan oleh model prediksi yang dibangun adalah 70,94 %. Dengan akurasi memprediksi *non churn* sebesar 71,31% dan akurasi memprediksi *churn* sebesar 58,02%. Hal ini membuktikan bahwa performansi model prediksi *churn* cukup baik. Tingkat akurasi prediksi *churn* yang lebih rendah membuktikan bahwa memprediksi pelanggan yang *churn* lebih sulit daripada memprediksi pelanggan yang tidak *churn*. Hal ini sesuai dengan kesimpulan penelitian [6].



Gambar 7. Kerangka Waktu Data Historis Ujicoba Sistem



Gambar 8. Model *Decison Tree* Ujicoba

**Tabel 1. Matriks Evaluasi Prediksi Ujicoba**

	NON-CHURN (Aktual)	CHURN (Aktual)
NON-CHURN (Hasil Prediksi)	6020	102
CHURN (Hasil Prediksi)	2422	141

## ANALISIS

Sistem yang telah dirancang melalui penelitian ini dapat memberikan nilai tambah pada CRM Telkomsel dalam pelaksanaan proses prediksi dan analisis *churn*. Sistem ini mempunyai kelebihan-kelebihan sebagai berikut:

- Dapat memaksimalkan pemanfaatan salah satu potensi yang dimiliki Telkomsel, yaitu data-data berukuran besar dalam menemukan pola perilaku *churn*.
- Dapat menghasilkan model prediksi *churn* yang dapat digunakan untuk memahami pola perilaku *churn* pelanggan.
- Menyediakan informasi yang mendukung perencanaan program retensi pelanggan.

Kelemahan-kelemahan dalam sistem yang dirancang adalah sebagai berikut:

- Tidak dapat mengendalikan kebenaran nilai suatu data
- Menghabiskan banyak waktu (*time consuming*) untuk *sourcing* dan *preprocessing* data.
- Hanya menggunakan teknik klasifikasi dengan algoritma *decision tree*, tidak membandingkan dengan teknik-teknik dan algoritma lain dalam pembangunan model prediksi.
- Model prediksi tidak mempertimbangkan variabel-variabel yang datanya tidak dimiliki Telkomsel. Seperti harga dan jenis produk yang ditawarkan kompetitor, jumlah *handset* yang dimiliki, pekerjaan, jenis kelamin dan lain-lain.

## KESIMPULAN

Setelah melakukan analisis kebutuhan, perancangan dan pengujian terhadap sistem yang dirancang, maka dapat diperoleh kesimpulan sebagai berikut:

- Sistem prediksi *churn* yang dirancang dapat meningkatkan kemampuan *analytical* CRM Telkomsel dengan cara mengolah dan menginterpretasikan data-data pelanggan yang dimiliki menjadi informasi dan pengetahuan yang mendukung perencanaan strategi retensi Telkomsel. Informasi dan pengetahuan ini berupa model perilaku *churn* pelanggan, matriks klasifikasi pelanggan dan hasil prediksi *churn* per pelanggan. Selain itu, sistem yang dirancang juga dapat mendukung aktivitas dan perencanaan operasional CRM Telkomsel dengan menyediakan proses-proses terstruktur yang akan menghasilkan informasi perilaku dan hasil prediksi *churn* pelanggan.

- Persoalan mengenai terpecah-pecahnya data pada masing-masing divisi Telkomsel diatasi dengan melakukan integrasi data pada proses awal analisis dan prediksi *churn* pada sistem
- *Missing value* dan nilai yang tidak masuk akal dapat diatasi dengan memasukkan fitur “pembersihan data” pada sistem.
- Aktivitas *sourcing* dan *preprocessing* data yang berukuran besar merupakan aktivitas kritis yang paling banyak menghabiskan waktu serta mempengaruhi kualitas data yang akan digunakan untuk membangun model.
- Dengan *churn rate* Telkomsel yang lebih kecil dari 10%, jumlah observasi *churn* jauh lebih sedikit daripada observasi non-*churn*, hal ini dapat membuat model prediksi hanya belajar untuk memprediksi pola perilaku pelanggan yang non *churn*. Untuk mengatasi hal ini, sistem dapat membedakan perlakuan antara observasi *churn* dan observasi non-*churn* dalam pembersihan data serta *sampling* dan pemisahan data.

## DAFTAR PUSTAKA

1. Richeldi, Marco and P. Alessandro, *Churn Analysis Case Study*, Telecom Italia Lab, Torino, 2002.
2. Mozer, Michael C., Richard Wolniewicz, David B. Grimes, E. Johnson, and Howard Kaushansky, *Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry*, IEEE Transactions on Neural Network, Volume 3, 2000.
3. Hutomo, Arry, *Desain Sistem Customer Relationship Management Difokuskan untuk Mengatasi Turnover Pelanggan*, Thesis Program Studi Teknik dan Manajemen Industri, Bandung, 2004.
4. Dyche, Jill, *The CRM Handbook*, Addison-Wesley, 2002.
5. Baragoïn, Corinne, Christian M. Andersen, Stephan Bayerl, Graham Bent, Jleun Lee, and Christoph Schommer, *Mining Your Own Business in Telecoms Using DB2 Intelligent Miner for Data*, International Business Machines, 2001.
6. Hadden, John, Ashutosh Tiwari, Rajkumar Roy, and Dymtr Ruta, *Churn Prediction using Complaints Data*, Transactions On Engineering, Computing And Technology, 13, 2006.
7. Euler, Timm, *Publishing Operational Models of Data Mining Case Studies*, Proceedings of the First International Workshop on Data mining Case Studies, Dortmund, 2005.
8. Han, Jiawei, Micheline Kamber, *Data Mining: Concepts and Techniques*, The Morgan Kaufmann Series, 2001.
9. Whiten L. Jeffrey, Lonnie D. Bentley, & K. C. Dittman, *Systems Analysis and Design Methods*, McGraw-Hill, Inc., New York, 2000.