

**SEGMENTASI DAN KLASIFIKASI PERILAKU PEMBAYARAN
PELANGGAN PADA PERUSAHAAN PENYEDIA LAYANAN
MULTIMEDIA DENGAN ALGORITMA
K-MEANS DAN C4.5**

TESIS

Diajukan Sebagai Salah Satu Syarat Untuk Menyelesaikan
Program Strata Dua (S2) Magister Komputer



SUSANTO HARIYANTO

371 210 1203

**PROGRAM STUDI TEKNIK INFORMATIKA
PROGRAM PASCA SARJANA (S2) MAGISTER KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER ERESHA
JAKARTA
2012**

PERSETUJUAN TESIS

Nama : Susanto Hariyanto
NPM : 371 210 1203
Konsentrasi : Rekayasa Perangkat Lunak
Judul tesis : Segmentasi dan Klasifikasi Perilaku Pembayaran Pelanggan pada Perusahaan Penyedia Layanan Multimedia dengan Algoritma K-Means dan C4.5

Telah disetujui untuk disidangkan pada Sidang Tesis pada Program Pasca Sarjana (S2) Magister Komputer, Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika dan Komputer Eresha.

Jakarta, September 2012

Pembimbing Utama

Pembimbing Pendamping

(Romi Satria Wahono, B.Eng., M.Eng)

(Dr. Rufman Iman Akbar E., SE, MM, M.Kom.)

Mengetahui :

Direktur

Progam Pasca Sarjana

(Dr. Rufman Iman Akbar E., SE, MM, M.Kom.)

LEMBAR PENGESAHAN

LEMBAR PERNYATAAN KEASLIAN TESIS

Nama : Susanto Hariyanto

NPM : 371 210 1203

Konsentrasi : Rekayasa Perangkat Lunak

Judul tesis : Segmentasi dan Klasifikasi Perilaku Pembayaran Pelanggan pada Perusahaan Penyedia Layanan Multimedia dengan Algoritma K-Means dan C4.5

Dengan ini saya menyatakan bahwa dalam Tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Jakarta, September 2012

(Susanto Hariyanto)

Susanto Hariyanto, 371 210 1203

Segmentasi dan Klasifikasi Perilaku Pembayaran Pelanggan pada Perusahaan Penyedia Layanan Multimedia dengan Algoritma K-Means dan C4.5; dibawah bimbingan Romi Satria Wahono, B.Eng., M.Eng dan Dr. Rufman Iman Akbar E., MM, M.Kom.

78 + xiii hal / 20 tabel / 29 gambar / 1 lampiran / 17 pustaka (2003 – 2011)

ABSTRAK

Sebagai salah satu perusahaan penyedia layanan jasa di Jakarta yang bergerak dalam penyedia layanan internet dan TV kabel. Perusahaan ini mengalami masalah dengan pelanggan yang menolak membayar setelah penggunaan jasa. Identifikasi pelanggan layak sulit dilakukan karena perusahaan penyedia layanan jasa tidak dapat melakukan verifikasi keuangan pelanggan. Selain itu semua pelanggan yang masuk dalam jaringan merupakan pelanggan yang potensial bagi perusahaan. Pada penelitian serupa, algoritma C4.5 mampu menghasilkan klasifikasi dengan tingkat akurasi yang baik dibandingkan dengan algoritma lain. Namun pemilihan atribut untuk pengolahan dengan algoritma C4.5 cukup sulit. Terlebih lagi perusahaan tidak memiliki data keuangan pelanggan yang menjadi salah satu atribut penting dalam mengukur kelayakan pelanggan. Oleh karena itu perilaku pembayaran pelanggan akan disegmentasi dengan *k-means* agar dapat dilihat tingkat potensial pelanggan yang akan digunakan sebagai atribut klasifikasi. Ekstraksi atribut potensial pelanggan dengan *k-means* ternyata mampu membantu meningkatkan akurasi dari klasifikasi algoritma C4.5. Hal ini terbukti dari peningkatan akurasi model dari 59.02 % menjadi 77.31% dan AUC dari 0.537 menjadi 0.836. Pengenalan potensial pelanggan juga dapat dijadikan acuan dalam proses promosi, retensi, dan penanggulangan pelanggan tidak layak.

Kata Kunci:

Kelayakan pelanggan, potensial pelanggan, algoritma C4.5, algoritma k-means

Susanto Hariyanto, 371 210 1203

Clustering and Classification Customer Payment Behaviour in Multimedia Company with K-Means and C4.5 Algorithm; under guidance of Romi Satria Wahono, B.Eng., M.Eng and Dr. Rufman Iman Akbar E., MM, M.Kom.

78 + xiii pages / 20 tables / 29 figures / 1 enclosure / 17 references (2003 – 2011)

ABSTRACT

A service provider company often have problem with insolvent customer. As a company that can not verify customer financial, a service provider company is having difficulty on identifying solvent customer. As one of the service provider company in Jakarta which offer internet connectivity and cable TV service, identifying solvent customer is important because, every customer inside coverage area is a potential customer to subscribe their product. In a related research, algorithm C4.5 become one of the favorite algorithm to use in classifying customer insolvency, and was able to produce high accuracy model, and easy to understand. The major problem of algorithm C4.5 is attribute selection. As noted that a service provider company cannot verify customer financial, they have no data about customer financial which is one important attribute in credit scoring. Therefore, customer payment behavior will be extracted as customer potential as one of classification attribute. Feature extraction with k-means was able to improve the accuracy of the algorithm C4.5. As noted by improve on accuracy from 59.02% to 77.31% and AUC from 0.537 to 0.836. The customer potential attribute can also be used as reference in the promotion, retention, and managing insolvent customer.

Keywords:

Customer insolvency, customer potential, algorithm C4.5, algorithm K-Means.

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan hidayahnya kepada penulis, sehingga tersusunlah tesis yang berjudul “Segmentasi dan Klasifikasi Perilaku Pembayaran Pelanggan pada Perusahaan Penyedia Layanan Multimedia dengan Algoritma K-Means dan C4.5”. Tesis tersebut melengkapi salah satu persyaratan yang diajukan dalam rangka menempuh ujian akhir untuk memperoleh gelar Magister Komputer (M.Kom.) pada Program Pasca Sarjana (S2), Program Studi Teknik Informatika di Sekolah Tinggi Manajemen Informatika dan Komputer Eresha

Penulis sungguh sangat menyadari, bahwa penulisan tesis ini tidak akan terwujud tanpa adanya dukungan dan bantuan dari berbagai pihak. Maka, dalam kesempatan ini penulis menghaturkan penghargaan dan ucapan terima kasih yang sebesar - besarnya kepada yang terhormat :

1. Bapak Ir.Damsiruddin Siregar, MMT, selaku Ketua STMIK Eresha.
2. Bapak Dr. Rufman Iman Akbar E., SE, MM, M.Kom, selaku Direktur Pasca Sarjana STMIK Eresha yang sekaligus menjadi dosen pembimbing pendamping yang telah banyak membantu memberikan saran-saran perbaikan dalam penulisan thesis ini.
3. Bapak Didik Setiyadi, M. Kom, selaku Puket II STMIK Eresha.
4. Bapak Bobby Reza, S.Kom., MM, selaku Puket III STMIK Eresha.
5. Bapak Romi Satria Wahono, B.Eng., M.Eng, selaku dosen pembimbing utama yang telah banyak membantu memberikan ide dan saran-saran dalam penulisan tesis ini.
6. Bapak/Ibu dosen STMIK Eresha yang telah memberikan ilmunya.
7. Rekan-rekan perjuangan Desiyanna, Edi, Rino, Dicky, dan Indah yang telah berbagi suka dan duka, membimbing dan menyematkan dalam mengerjakan thesis ini hingga perjuangan terkahir.
8. Rekan-rekan mahasiswa angkatan 37 (pusat) STMIK Eresha yang telah berjuang bersama dalam perkuliahan.

Akhir kata mohon maaf atas kekeliruan dan kesalahan yang ada dalam tesis ini, baik yang disengaja maupun tidak disengaja dan berharap semoga tesis ini dapat memberikan manfaat bagi khasanah pengetahuan teknologi informasi di Indonesia.

Penulis

DAFTAR ISI

| | Hal. |
|--|------|
| PERSETUJUAN TESIS | i |
| LEMBAR PENGESAHAN | ii |
| LEMBAR PERNYATAAN KEASLIAN TESIS | iii |
| ABSTRAK..... | iv |
| ABSTRACT..... | v |
| KATA PENGANTAR | vi |
| DAFTAR ISI..... | viii |
| DAFTAR TABEL..... | xi |
| DAFTAR GAMBAR..... | xii |
| DAFTAR LAMPIRAN..... | xiii |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Permasalahan Penelitian..... | 3 |
| 1.2.1 Identifikasi Masalah | 3 |
| 1.2.2 Ruang Lingkup Masalah..... | 4 |
| 1.2.3 Rumusan Masalah | 4 |
| 1.3 Tujuan dan Manfaat Penelitian..... | 5 |
| 1.3.1 Tujuan Penelitian..... | 5 |
| 1.3.2 Manfaat Penelitian..... | 5 |
| 1.4 Sistematika Penulisan..... | 6 |
| BAB II LANDASAN TEORI DAN KERANGKA PEMIKIRAN | 8 |
| 2.1 Tinjauan Pustaka | 8 |
| 2.1.1 Model penelitian Daskalaki..... | 8 |
| 2.1.2 Model Penelitian Pinheiro | 10 |
| 2.1.3 Model Penelitian Ali..... | 11 |

| | | |
|-----------------------------------|---|----|
| 2.1.4 | Rangkuman penelitian terkait | 12 |
| 2.2 | Landasan Teori | 13 |
| 2.2.1 | Segmentasi Pelanggan | 13 |
| 2.2.2 | Kelayakan Pelanggan | 15 |
| 2.2.3 | Algoritma Klasifikasi | 17 |
| 2.2.4 | Algoritma Segmentasi | 28 |
| 2.2.5 | Evaluasi dan Pengujian Algoritma Klasifikasi | 35 |
| 2.3 | Kerangka Pemikiran | 36 |
| BAB III METODE PENELITIAN | | 38 |
| 3.1 | Analisa Kebutuhan | 38 |
| 3.2 | Perancangan Penelitian | 40 |
| 3.3 | Teknik Analisis | 41 |
| 3.3.1 | Pengumpulan Data | 41 |
| 3.3.2 | Pengolahan Data Awal | 42 |
| 3.3.3 | Model yang Diusulkan | 46 |
| 3.3.4 | Eksperimen dan Pengujian Model | 48 |
| 3.3.5 | Evaluasi dan Hasil | 53 |
| BAB IV HASIL DAN PEMBAHASAN | | 54 |
| 4.1 | Hasil | 54 |
| 4.1.1 | Segmentasi Potensial Pelanggan | 54 |
| 4.1.2 | Klasifikasi Kelayakan Pelanggan | 59 |
| 4.2 | Pembahasan | 68 |
| 4.2.1 | Evaluasi model <i>Confusion Matrix</i> | 69 |
| 4.2.2 | Kurva ROC (Receiver Operating Characteristic) | 71 |
| 4.3 | Implikasi Penelitian | 73 |
| BAB V KESIMPULAN DAN SARAN | | 75 |
| 5.1 | Kesimpulan | 75 |

| | | |
|-----|----------------------------|----|
| 5.2 | Saran..... | 76 |
| | DAFTAR PUSTAKA..... | 77 |
| | DAFTAR RIWAYAT HIDUP | 79 |
| | LAMPIRAN-LAMPIRAN | 80 |

DAFTAR TABEL

| | Hal. |
|--|------|
| Tabel 2.1 Perbandingan Penelitian Serupa | 13 |
| Tabel 2.2 Data pembelian komputer | 26 |
| Tabel 2.3 Data yang akan disegmentasi | 33 |
| Tabel 2.4 Nilai titik tengah untuk setiap segmentasi | 34 |
| Tabel 2.5 <i>Confusion matrix</i> dengan hasil yang baik | 35 |
| Tabel 3.1 Data awal yang belum diolah (data untuk rate, balance, pay dan age terdiri dari 6 bulan) | 39 |
| Tabel 3.2 Data setelah transformasi | 45 |
| Tabel 3.3 Atribut yang digunakan | 45 |
| Tabel 3.4 Spesifikasi komputer yang digunakan | 48 |
| Tabel 4.1 Contoh data untuk segmentasi pelanggan | 54 |
| Tabel 4.2 Data yang akan disegmentasi dengan <i>k-means</i> | 55 |
| Tabel 4.3 Hasil perhitungan titik euclidean pertama | 56 |
| Tabel 4.4 Nilai tengah baru untuk setiap cluster(Tahap 1) | 56 |
| Tabel 4.5 Tabel nilai <i>euclidean</i> pada tahap ke 2 | 57 |
| Tabel 4.6 Tabel Statistik Keseluruhan Data Untuk Segmentasi | 57 |
| Tabel 4.7 Nilai Titik Tengah Setiap Segmen | 58 |
| Tabel 4.8 Data contoh untuk perhitungan klasifikasi kelayakan pelanggan | 60 |
| Tabel 4.9 Nilai Pengujian Indikator | 62 |
| Tabel 4.10 Nilai akurasi pengujian indikator <i>minimal gain</i> dan <i>pruning</i> | 70 |
| Tabel 4.11 <i>Confusion matrix</i> untuk dataset yang tidak disegmentasi | 70 |
| Tabel 4.12 <i>Confusion matrix</i> untuk data setelah segmentasi | 71 |
| Tabel 4.13 Nilai AUC pengujian indikator <i>minimal gain</i> dan <i>pruning</i> | 71 |
| Tabel 4.14 Perbandingan model C4.5 sebelum dan sesudah segmentasi | 73 |
| Tabel 4.15 Segmen pelanggan yang terbentuk | 74 |

DAFTAR GAMBAR

| | Hal. |
|--|------|
| Gambar 2.1 Model yang diusulkan oleh Daskalaki. | 9 |
| Gambar 2.2 Model yang diusulkan oleh Pinheiro | 10 |
| Gambar 2.3 Model yang diusulkan oleh Ali | 12 |
| Gambar 2.4 Contoh pohon keputusan..... | 19 |
| Gambar 2.5 Pembuatan cabang pohon keputusan | 20 |
| Gambar 2.8 Pohon keputusan yang dipangkas dengan <i>postprunning</i> | 24 |
| Gambar 2.9 Contoh repetisi pohon keputusan pada atribut A1 | 24 |
| Gambar 2.10 Perulangan dalam pohon keputusan..... | 25 |
| Gambar 2.6 keputusan setelah dicabangkan dengan <i>age</i> | 27 |
| Gambar 2.7 Pemilihan titik perpecahan dalam atribut..... | 28 |
| Gambar 2.11 Pengambilan titik tengah setiap segmen | 32 |
| Gambar 2.12 Pusat segmen bergeser, dan objek yang ada dalam segmentasi disesuaikan hingga tidak ada perubahan | 32 |
| Gambar 2.13 Data acak yang akan digunakan sebagai data awal segmen | 33 |
| Gambar 2.14 Kumpulan data awal untuk setiap segmen | 34 |
| Gambar 2.15 Kerangka Pemikiran..... | 37 |
| Gambar 3.1 Detail model yang diusulkan..... | 47 |
| Gambar 3.2 Import data dari dokumen CSV. | 49 |
| Gambar 3.3 Tampilan attribut dari data yang akan diolah..... | 49 |
| Gambar 3.4 Parameter untuk segmentasi pelanggan. | 50 |
| Gambar 3.5 Relasi antar proses import data dengan proses Clustering..... | 50 |
| Gambar 3.6 <i>Centroid Table</i> dari <i>Cluster Model</i> yang terbentuk | 51 |
| Gambar 3.7 Building Block untuk C4.5 | 52 |
| Gambar 3.8 Parameter untuk Proses <i>Decision Tree</i> | 52 |
| Gambar 3.9 Relasi untuk mengklasifikasikan kelayakan pelanggan. | 53 |
| Gambar 4.1 Pohon keputusan yang terbentuk untuk segmen pertama | 63 |
| Gambar 4.2 Pohon Keputusan untuk segmen kedua..... | 64 |
| Gambar 4.3 Pohon Keputusan untuk segmen ketiga. | 64 |
| Gambar 4.4 Pohon Keputusan untuk segmen keempat..... | 65 |
| Gambar 4.5 Pohon Keputusan untuk segmen kelima | 66 |

DAFTAR LAMPIRAN

Hal.

No table of figures entries found.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu perusahaan pelayanan jasa yang bergerak dalam layanan tv kabel dan internet di Jakarta. Memberikan layanan kepada pelanggan dengan infrastruktur kabel *fiber optic* ke areal perumahan. Pelanggan yang masuk kedalam jaringan pelayanan, akan dihubungi dan ditawarkan promosi penggunaan jasa. Penawaran diberikan kepada setiap pelanggan yang masuk dalam jangkauan tanpa verifikasi detail terlebih dahulu. Jika pelanggan yang sudah berlangganan terlambat atau menolak membayar, perusahaan tetap memberikan jasa dengan pembatasan tertentu selama 3 bulan (pembatasan *channel* untuk tv kabel, dan pembatasan kecepatan untuk internet) dengan harapan pelanggan dapat terus menggunakan jasa.

Perusahaan penyedia layanan multimedia sering mengalami masalah dengan pelanggan yang menolak melakukan pembayaran atas jasa yang sudah digunakan (Daskalaki, Kopanas, Goudara, & Avouris, 2003). Berbeda dengan bank atau perusahaan yang menawarkan pinjaman dana, perusahaan jasa yang pascabayar seringkali memberikan jasanya kepada pelanggan tanpa melalui proses verifikasi yang detail, sehingga sulit untuk mengetahui apakah pelanggan yang berlangganan merupakan pelanggan yang layak (Pinheiro, Evsukoff, & Ebecken, 2006). Oleh karena itu terdapat pelanggan yang menolak melakukan pembayaran sehingga menimbulkan hutang dan mengurangi keuntungan.

Perusahaan jasa memiliki regulasi untuk tetap melanjutkan layanan dalam periode tertentu walaupun pelanggan menolak membayar (Pinheiro, Evsukoff, & Ebecken, 2006). Meskipun adanya penalti yang dikenakan, namun tetap menjadi masalah. Deteksi dan pencegahan perilaku pelanggan yang menolak membayar ini adalah salah satu tujuan yang penting untuk dicapai dalam industri.

Dalam klasifikasi kelayakan, salah satu atribut yang sangat berpengaruh adalah keuangan pelanggan. Namun perusahaan penyedia layanan multimedia tidak memiliki data detail mengenai keuangan pelanggan (Daskalaki, Kopanas, Goudara, & Avouris, 2003). Oleh karena itu data pembayaran pelanggan dapat disegmentasi untuk melihat potensial pelanggan dan membantu perusahaan dalam

melakukan tindakan pencegahan yang tepat sesuai dengan segmen pelanggan (Pinheiro, Evsukoff, & Ebecken, 2006). Dengan demikian, perusahaan dapat mengambil tindakan yang sesuai dengan kelompok pelanggan.

Metode *data mining* saat ini sudah banyak digunakan untuk meningkatkan bisnis dan meningkatkan keuntungan dalam kompetisi dengan perusahaan pesaing. *Data mining* dapat mengolah data dalam jumlah besar dan menghasilkan pengetahuan yang dijadikan acuan dalam keputusan bisnis (Thiruvadi & Patel, 2011). *Data mining* dapat digunakan untuk klasifikasi dan segmentasi. Teknik segmentasi merupakan teknik yang memecah data kedalam kelompok-kelompok yang homogen.. Sedangkan teknik klasifikasi bertujuan untuk melihat perilaku pelanggan yang tidak layak terutama saat menjelang periode pembayaran.

Data mining sudah banyak digunakan untuk masalah perilaku pelanggan, (Liao & Triantaphyllou, 2007) menunjukkan banyak penelitian mengenai data mining, dimana penelitian yang melibatkan pelanggan menjadi salah satu kategori besar. (Thiruvadi & Patel, 2011) melakukan survei terhadap penerapan *data mining* dalam deteksi dan pencegahan kecurangan dimana disebutkan bahwa pelanggan yang memakai jasa dan menolak membayar juga tergolong dalam kecurangan.

Pelanggan perusahaan penyedia jasa yang terlambat atau menolak membayar memberikan dampak negatif bagi perusahaan (Daskalaki, Kopanas, Goudara, & Avouris, 2003), oleh karena itu banyak penelitian yang menggunakan *machine learning* untuk memprediksi pelanggan yang tidak baik, serta melakukan tindakan pencegahan. C4.5 merupakan salah satu algoritma yang populer dengan tingkat akurasi yang tinggi pada penelitian serupa. Namun jika digunakan untuk data dengan atribut yang banyak, serta kurang relevan dan redundan dapat mengurangi akurasi algoritma C4.5 (Liu & Schumann, 2005). Oleh karena itu perlu dilakukan pemilihan atau pengelompokkan atribut untuk mengurangi atribut yang redundan.

Perilaku pelanggan perlu dipelajari agar kebijakan yang lebih sesuai dapat diambil, oleh karena itu perilaku pelanggan akan disegmentasi dengan K-Means. Selain itu segmentasi perilaku ini juga membantu mengurangi atribut pembayaran yang cenderung redundan. Pemrosesan data awal dengan *k-means* dapat

meningkatkan akurasi dari algoritma klasifikasi (Ali, Sulaiman, Mustapha, & Mustapha, 2009). Selain itu dengan mempelajari perilaku pelanggan, kebijakan yang dibuat dapat lebih spesifik untuk mempertahankan serta meningkatkan relasi dengan pelanggan.

(Daskalaki, Kopanas, Goudara, & Avouris, 2003) menerapkan data mining untuk membantu memprediksi pelanggan yang menolak membayar pada perusahaan telekomunikasi. (Pinheiro, Evsukoff, & Ebecken, 2006) menggunakan *self organizing map* untuk segmentasi pelanggan dan diprediksi dengan *neural network* untuk menentukan kelayakan pelanggan. (Prasad & Malik, 2011) melakukan segmentasi pelanggan untuk mendapatkan pola perilaku pelanggan.

Pada penelitian ini, pelanggan akan disegmentasi dengan algoritma *k-means* menurut perilaku pembayaran pelanggan agar dapat diukur tingkat potensial pelanggan. Setiap segmen pelanggan akan di klasifikasikan menurut kelayakan pelanggan dengan algoritma C4.5. Dengan begitu akurasi algoritma C4.5 akan lebih baik serta lebih cocok untuk diterapkan menurut tingkat potensial pelanggan.

1.2 Permasalahan Penelitian

1.2.1 Identifikasi Masalah

Dalam penelitian ini, perusahaan penyedia layanan jasa mengalami kesulitan dalam mengidentifikasi pelanggan yang layak, untuk mengidentifikasi kelayakan pelanggan, untuk dapat mengidentifikasikan kelayakan pelanggan, algoritma data mining dapat digunakan untuk mempelajari data yang dimiliki, namun ada beberapa masalah yang dihadapi dalam proses klasifikasikan pelanggan yang tidak layak:

- a. Dalam penelitian serupa, algoritma C4.5 merupakan algoritma yang paling banyak digunakan, karena memiliki hasil akurasi yang tinggi, serta model yang terbentuk mudah dibaca. Namun akurasi C4.5 sangat ditentukan oleh pemilihan attribut yang tepat, relevan, serta tidak terlalu banyak.
- b. Perusahaan penyedia layanan jasa tidak memiliki data mendetail mengenai tingkat keuangan pelanggan. Sementara data keuangan pelanggan merupakan

attribut yang penting dalam menentukan kelayakan pelanggan (Daskalaki, Kopanas, Goudara, & Avouris, 2003).

- c. Setelah mengetahui kelayakan pelanggan, perusahaan perlu mengambil tindakan yang sesuai untuk dapat menanggulangi kelayakan pelanggan. Tindakan yang diambil harus disesuaikan dengan tingkat potensial pelanggan.

1.2.2 Ruang Lingkup Masalah

Penelitian ini akan mengklasifikasikan kelayakan pelanggan pada salah satu perusahaan layanan jasa internet dan TV kabel di Jakarta. Proses pembayaran ditagihkan setiap bulan setelah pemakaian jasa. Dan pelanggan yang tidak membayar tagihan pada waktunya tetap dapat menikmati layanan selama tiga bulan dengan penalti tertentu. Oleh karena itu perusahaan ingin mengenali pelanggan yang tidak layak agar dapat ditanggulangi secara langsung tanpa menunggu 3 bulan.

Data penelitian akan diambil dari data pembayaran pelanggan, serta data lain yang dapat dikumpulkan seperti keluhan pelanggan dan jasa yang digunakan. Data dikumpulkan hingga akhir tahun 2011 dan hanya data pelanggan yang berusia lebih dari 6 bulan yang digunakan dalam penelitian ini.

1.2.3 Rumusan Masalah

Agar perusahaan dapat memiliki model yang baik dalam memprediksi kelayakan pelanggan, data pelanggan yang sudah ada akan dikumpulkan dan akan diolah dengan algoritma C4.5. Untuk membantu mengurangi jumlah attribut yang akan diolah dengan algoritma C4.5 dalam klasifikasi pelanggan. Attribut pembayaran pelanggan akan dikelompokkan dengan algoritma *K-Means* sehingga dapat diketahui tingkat potensial pelanggan.

Dengan tambahan attribut potensial pelanggan, diharapkan klasifikasi kelayakan pelanggan akan lebih akurat. Selain itu perusahaan juga dapat menyesuaikan tindakan pencegahan dan penanggulangan serta dapat memberikan promosi yang sesuai dengan tingkat potensial pelanggan. Seberapa meningkatnya akurasi algoritma C4.5 untuk klasifikasi kelayakan pelanggan yang akan diterapkan pada data yang sudah disegmentasi dengan algoritma *K-Means*?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah menciptakan model yang baik untuk perusahaan agar dapat mengenali potensial pelanggan berdasarkan perilaku pembayaran yang akan disegmentasi dengan algoritma *K-Means*, dan menerapkan algoritma C4.5 pada setiap segmentasi pelanggan agar hasil klasifikasi menjadi lebih akurat. Dengan demikian perusahaan dapat mengenal kelayakan pelanggan, serta mengambil tindakan yang sesuai dengan potensial pelanggan.

1.3.2 Manfaat Penelitian

Dengan adanya segmentasi pelanggan, perusahaan dapat membedakan kelas pelanggan yang berpotensi, sedang dan tidak berpotensi. Perusahaan dapat lebih mengenal pelanggan yang potensial sehingga membantu dalam pembuatan kebijakan yang dapat meningkatkan relasi antara pelanggan dengan perusahaan. Selain itu perusahaan juga dapat mengambil tindakan spesifik dalam menangani pelanggan yang menolak membayar sesuai dengan kelas mereka masing masing.

- a. Dengan prediksi kelayakan pelanggan, perusahaan juga dapat menilai pelanggan yang dimiliki. Perusahaan dapat mengenal pelanggan yang baik untuk digunakan sebagai objek promosi kedepan. Selain itu perusahaan juga dapat menggunakan nilai kelayakan pelanggan untuk membangun kebijakan strategis dalam meningkatkan loyalitas pelanggan.
- b. Dengan terprediksinya pelanggan yang tidak baik, maka perusahaan dapat dengan langsung mengambil tindakan sesuai dengan kelas pelanggan. Sehingga perusahaan dapat meminimalisir dampak negatif yang muncul. Walaupun perusahaan tetap harus memberikan layanannya dengan penalti bagi pelanggan yang tidak baik, namun usaha pendekatan dapat dikurangi atau bahkan tidak perlu dilakukan jika pelanggan memang sudah masuk dalam kelas yang tidak berpotensi. Pendekatan terhadap pelanggan yang potensialpun dapat dilakukan dengan lebih efektif.
- c. Dapat menerapkan teknik clustering *K-Means* untuk mengelompokkan beberapa atribut yang kurang berpengaruh dalam proses klasifikasi menjadi satu atribut baru. Dengan mengelompokkan pola pembayaran dan penggunaan

jasa pelanggan, dapat diketahui potensial pelanggan. Dengan atribut potensial pelanggan, atribut-atribut penyusunnya dapat dihilangkan sehingga dapat mengurangi dimensi data yang diolah oleh algoritma C4.5. Sehingga model yang terbentuk dapat lebih mudah dipahami, serta memiliki akurasi yang lebih baik.

1.4 Sistematika Penulisan

Sistematika penulisan pada penelitian ini adalah sebagai berikut:

Bab I Pendahuluan

Bab I berisi penjelasan tentang latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian dan kontribusi serta sistematika penulisan.

Bab II Landasan Teori dan Kerangka Pemikiran

Bab II berisi penjelasan tentang landasan teori yang digunakan dalam penelitian dan kerangka pemikiran. Diuraikan pula tentang tinjauan pustaka yang merupakan penjelasan tentang hasil-hasil penelitian lainnya yang berkaitan dengan penelitian yang dilakukan. Landasan teori merupakan suatu penjelasan tentang sumber acuan terbaru dari pustaka primer seperti buku, artikel, jurnal, dan tulisan asli lainnya untuk mengetahui perkembangan penelitian yang relevan dengan judul atau tema penelitian yang dilakukan dan juga sebagai arahan dalam memecahkan masalah yang diteliti. Dalam bab ini juga diuraikan tentang kerangka pemikiran yang merupakan penjelasan tentang kerangka berpikir untuk memecahkan masalah yang sedang diteliti, termasuk menguraikan objek penelitian.

Bab III Metodologi Penelitian

Bab III ini merupakan penjelasan tentang metode penelitian, penentuan masalah, penentuan model dan metode, juga penjelasan bagaimana cara pengumpulan data, pengelompokkan data pelanggan dengan k-Means, serta mengklasifikasikan data pelanggan dengan C4.5, diuraikan pula cara evaluasi dan validasi hasil penerapan, komparasi metode, metode analisis data, metode pengukuran penelitian, penerapan model dan metode pada masalah penelitian, pengembangan prediksi kelayakan pelanggan yang menerapkan model dan metode, analisa kebutuhan, perancangan model, konstruksi model dan pengujian model serta komparasi metode.