

**PREDIKSI LOYALITAS PELANGGAN PADA PERUSAHAAN
PENYEDIA LAYANAN MULTIMEDIA DENGAN
SEGMENTASI K-MEANS DAN ALGORITMA C4.5**

TESIS



PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)

PROGRAM PASCASARJANA

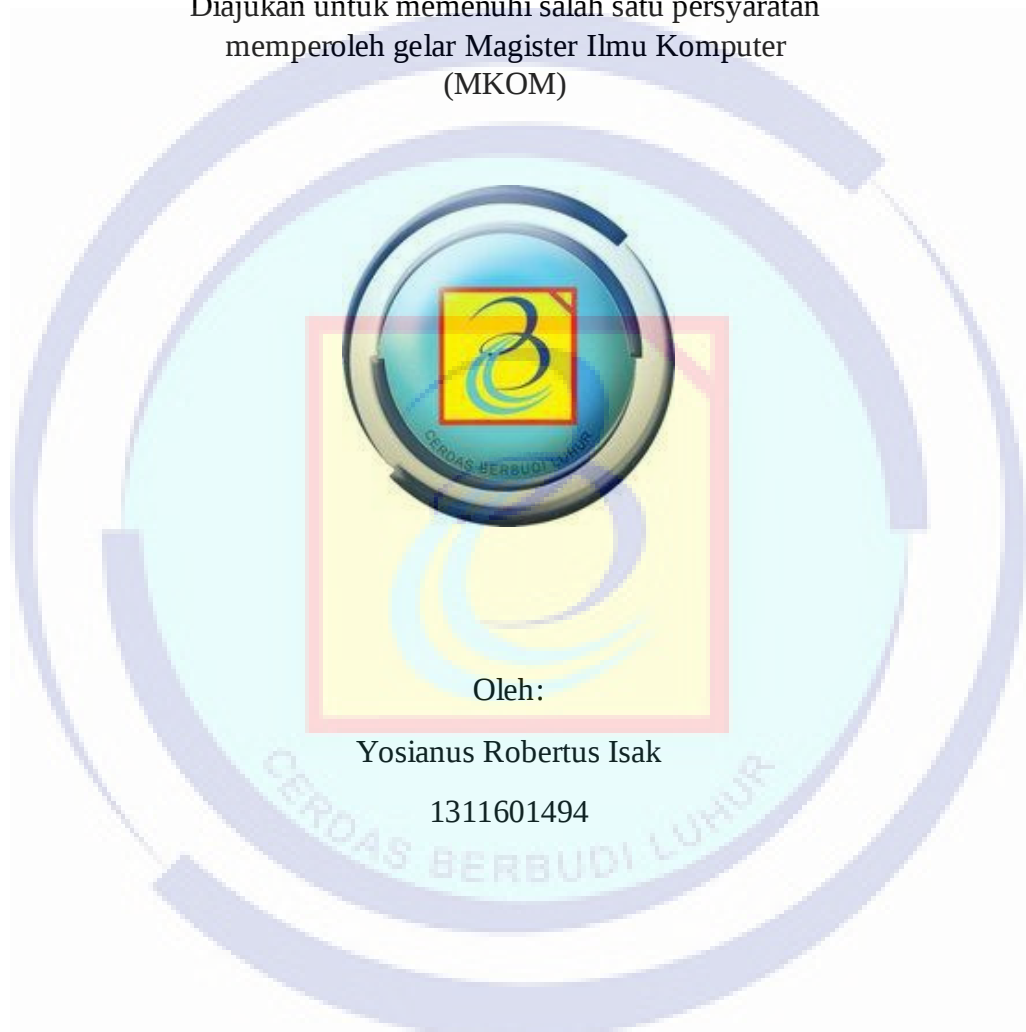
UNIVERSITAS BUDI LUHUR

JAKARTA

2016

**PREDIKSI LOYALITAS PELANGGAN PADA PERUSAHAAN
PENYEDIA LAYANAN MULTIMEDIA DENGAN
SEGMENTASI K-MEANS DAN ALGORITMA C4.5**

Diajukan untuk memenuhi salah satu persyaratan
memperoleh gelar Magister Ilmu Komputer
(MKOM)



Oleh:

Yosianus Robertus Isak

1311601494

PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)

PROGRAM PASCA SARJANA

UNIVERSITAS BUDI LUHUR

JAKARTA

2016



PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama Mahasiswa : Yosianus Robertus Isak
Nomor Induk Mahasiswa : 1311601494
Program Studi : Magister Ilmu Komputer
Konsentrasi : Teknologi Sistem Informasi
Fakultas/Program : Pascasarjana

Menyatakan bahwa Tesis yang berjudul:

PREDIKSI LOYALITAS PELANGGAN PADA PERUSAHAAN PENYEDIA LAYANAN MULTIMEDIA DENGAN SEGMENTASI K-MEANS DAN ALGORITMA C4.5

1. Merupakan hasil karya tulis ilmiah sendiri dan bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik oleh pihak lain,
2. Saya ijin untuk dikelola oleh Universitas Budi Luhur sesuai dengan norma hukum dan etika yang berlaku.

Pernyataan ini saya buat dengan penuh tanggung jawab dan saya bersedia menerima konsekuensi apapun sesuai aturan yang berlaku apabila di kemudian hari pernyataan ini tidak benar.

Jakarta, 03 Februari 2016



(Yosianus Robertus Isak)



PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PENGESAHAN

Nama Mahasiswa : Yosianus Robertus Isak
Nomor Induk Mahasiswa : 1311601494
Konsentrasi : Teknologi Sistem Informasi
Jenjang Studi : Magister Ilmu Komputer
Judul Proposal Tesis : PREDIKSI LOYALITAS PELANGGAN PADA
PERUSAHAAN PENYEDIA LAYANAN
MULTIMEDIA DENGAN SEGMENTASI K-
MEANS DAN ALGORITMA C4.5

Jakarta, 03 Februari 2016

Tim Penguji:

Tanda Tangan:

Ketua,
(Dr. Ir. Nazori AZ., MT)

Anggota,
(Arief Wibowo., M.Kom)

Pembimbing Utama,
(Prof. DR. Moedjiono, M.Sc)

Ketua Program Studi

(Dr. Ir. Nazori AZ., MT)

ABSTRAK

Saat ini, Perusahaan penyedia layanan jasa multimedia seperti internet dan *tv cable* merupakan perusahaan yang berkembang, terbukti dari pengguna internet yang semakin meningkat setiap tahunnya. Hal ini memunculkan peluang bagi beberapa pihak untuk membuat banyak perusahaan penyedia jasa layanan multimedia sehingga pelanggan memiliki banyak pilihan dalam memilih perusahaan. Mempertahankan pelanggan merupakan hal yang penting karena biaya untuk mempertahankan pelanggan jauh lebih murah daripada mendapatkan pelanggan baru. Sehingga sangat penting bagi perusahaan untuk mengetahui loyalitas pelanggan sehingga perusahaan juga dapat membuat proyeksi pendapatan sebagai acuan dalam rencana pembangunan perusahaan. Salah satu algoritma data mining yang cukup baik dalam menentukan klasifikasi loyalitas pelanggan adalah algoritma C4.5 karena mampu menghasilkan model yang mudah dimengerti dengan akurasi yang tinggi. Akurasi model C4.5 sangat tergantung pada atribut pembentuknya, oleh karena itu pengolahan awal data sangat penting untuk menghasilkan model sederhana yang tingkat akurasinya tinggi. Algoritma segmentasi *k-means* akan diterapkan untuk melakukan segmentasi terhadap beberapa atribut sehingga dapat digunakan sebagai atribut yang lebih bernilai dan memperkecil jumlah atribut yang akan digunakan guna meningkatkan akurasi dari algoritma klasifikasi C4.5. Dengan 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 69.23 % menjadi 79.33% dan AUC dari 0.723 menjadi 0.831. Pengenalan potensial pelanggan juga dapat dijadikan acuan dalam proses promosi, retensi, dan penanggulangan pelanggan tidak loyal.

Kata kunci: *Loyalitas pelanggan, algoritma C4.5, algoritma K- Means, prediksi, perusahaan multimedia, datamining*

ABSTRACT

Nowdays, multimedia service provider company like internet and tv cable is the company that grows up, that proves with the increase of internet user every years. This gives the opportunity for some people to make a lot of multimedia service provider so customer has a lot of provider choices. To keep the customer is the important thing because cost of keeping is cheaper than the cost to get a new one. It's so important for company to know the customer loyalty so company also can make projection of income as a benchmark to plan company to develop the company to be better. One of a lot datamining algoritms that is good enough to clasify customer loyalty is C4.5 algoritm. Because that can makes the model that easily to understand with the high accuracy. Accuracy of C4.5 model is depended by the attribut those are used, therefore data preprocessing is so important to produces the simple model with high accuracy. With feature extraction with k-means was able to improve the accuracy of the algorithm C4.5. As noted by improve on accuracy from 69.23% to 79.33% and AUC from 0.723 to 0.831. The customer potential attribute can alsobe used asreference in the promotion, retention, and managing non loyal customer.

Keywords: Customer loyalty, C4.5 Algorithm, K-means Algorithm, prediction, multimedia company, datamining

LEMBAR PERSEMBAHAN

Tidak ada hal apapun di dunia yang tidak dapat di pelajari, jika kita mau mengusahakannya.

(Yosianus Robertus Isak)

Dengan mengucap puji syukur kepada Tuhan Yang Maha Esa, tesis ini kupersembahkan untuk:

- 1 Bapak Agustinus Isak dan Ibu Marta, tercinta yang telah membesarkan aku dan selalu membimbing, mendukung, memotivasi, memberi apa yang terbaik bagiku serta selalu mendoakan aku untuk meraih kesuksesanku.
- 2 Teman – teman dari keluarga besar Universitas Buddhi Dharma, yang selalu memberi dukungan dan memberikan semangat.

KATA PENGANTAR

Dengan memanjatkan puji dan syukur kehadiran Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan berkahnya kepada penulis, sehingga penulis bisa menyelesaikan tesis yang berjudul *Prediksi Loyalitas Pelanggan Pada Perusahaan Penyedia Layanan Multimedia Dengan Segmentasi K – Means Dan Algoritma C4.5*. Tujuan dari penulisan penelitian ini adalah sebagai salah satu syarat untuk menyelesaikan Program Studi Magister Ilmu Komputer, Universitas Budi Luhur Jakarta.

Penulis sungguh sangat menyadari, bahwa dalam menyusun penelitian 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:

1. Bapak Prof. DR. Moedjiono, M.Sc selaku dosen pembimbing tesis yang telah membimbing dan memotivasi penulis dalam mengerjakan proposal penelitian proposal tesis ini.
2. Kedua orangtua, Bpk. Agustinus Isak dan Ibu Marta terima kasih atas dukungan doa, semangat dan kasih sayang yang senantiasa diberikan kepada penulis.
3. Keluarga Besar terima kasih atas dukungan doa, semangat dan kasih sayang yang senantiasa diberikan kepada penulis.
4. Bapak dan Ibu Dosen pengampu mata kuliah di Program Studi Magister Ilmu Komputer Universitas Budi Luhur yang telah dengan sabar memberikan ilmu pengetahuan, pencerahan, dan bimbingan dalam menuntut ilmu.
5. Rekan-rekan mahasiswa MKOM Universitas Budi Luhur angkatan 2013.
6. Teman-teman dari keluarga besar Universitas Buddhi Dharma dan semua pihak yang tidak dapat disebutkan satu persatu.

Penulis menyadari, sebagai manusia biasa yang tidak luput dari kekurangan dan kelemahan serta masih jauh dari kesempurnaan dalam penyusunan penelitian

ini, oleh karena itu penulis mengharapkan kritik dan saran yang bersifat membangun demi kesempurnaan penelitian nantinya. Semoga penelitian ini masih dapat memberikan manfaat bagi para pembaca dan juga para penulis penelitian selanjutnya.

Penulis



DAFTAR ISI

	Halaman
ABSTRAK.....	i
ABSTRACT.....	ii
LEMBAR PERSEMBAHAN	iii
KATA PENGANTAR	iv
DAFTAR ISI	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	3
1.2.1 Identifikasi Masalah	3
1.2.2 Batasan Masalah	3
1.2.3 Rumusan Masalah.....	4
1.3 Tujuan dan Manfaat Penelitian	4
1.3.1 Tujuan Penelitian	4
1.3.2 Manfaat Penelitian	4
1.4 Tata Urut Penulisan.....	5
BAB II LANDASAN TEORI DAN KERANGKA KONSEP	7
2.1 Tinjauan Pustaka	7
2.1.1 Data Mining	7
2.1.2 Loyalitas Pelanggan.....	8
2.1.3 Algoritma Klasifikasi C4.5	10
2.1.4 Algoritma Segmentasi <i>K-Means</i>	19
2.1.5 Evaluasi dan Pengujian Algoritma Klasifikasi	26
2.2 Tinjauan Objek Penelitian	27
2.2.1 Prediksi Loyalitas Pelanggan	27
2.2.2 Dataset Perusahaan Penyedia Jasa Layanan Multimedia	27

2.3	Tinjauan Studi	28
2.3.1	Model Penelitian Aishwarya ^([Aishwarya 2005], 225)	28
2.3.2	Model Penelitian John ^([John 2006], 1)	29
2.3.3	Model Penelitian Ali ^([Ali 2009], 1256)	30
2.3.4	Rangkuman Penelitian Terkait	31
2.4	Kerangka Pemikiran.....	32
2.5	Hipotesis	35
BAB III METODE PENELITIAN		36
3.1	Jenis Penelitian	36
3.2	Metode Pengumpulan Data	38
3.2.1	Metode Pengolahan Data Awal	38
3.2.2	Studi Kepustakaan	38
3.3	Langkah – Langkah Penelitian	38
3.4	Design Penelitian	40
3.4.1	Pengumpulan Data.....	40
3.4.2	Pengolahan Data	41
3.4.3	Model/ Metode yang Diusulkan	46
3.4.4	Eksperimen dan Pengujian Model.....	48
3.4.5	Evaluasi dan Validasi Hasil	56
3.5	Jadwal Penelitian.....	57
BAB IV PEMBAHASAN HASIL PENELITIAN		59
4.1	Hasil Penelitian	59
4.1.1	Segmentasi Potensial Pelanggan	59
4.1.2	Klasifikasi Loyalitas Pelanggan	66
4.2	Evaluasi dan Validasi.....	74
4.2.1	Evaluasi Model Confusion Matrix	75
4.2.2	Kurva ROC(Receiver Operating Characteristic)	77
4.3	Penerapan Algoritma Terpilih	79
4.4	Implikasi Penelitian	82
4.5	Rencana Implementasi.....	83
BAB V PENUTUP		85
5.1	Kesimpulan.....	85

5.2 Saran	86
DAFTAR PUSTAKA	87
LAMPIRAN	89
RIWAYAT HIDUP SINGKAT	111



DAFTAR GAMBAR

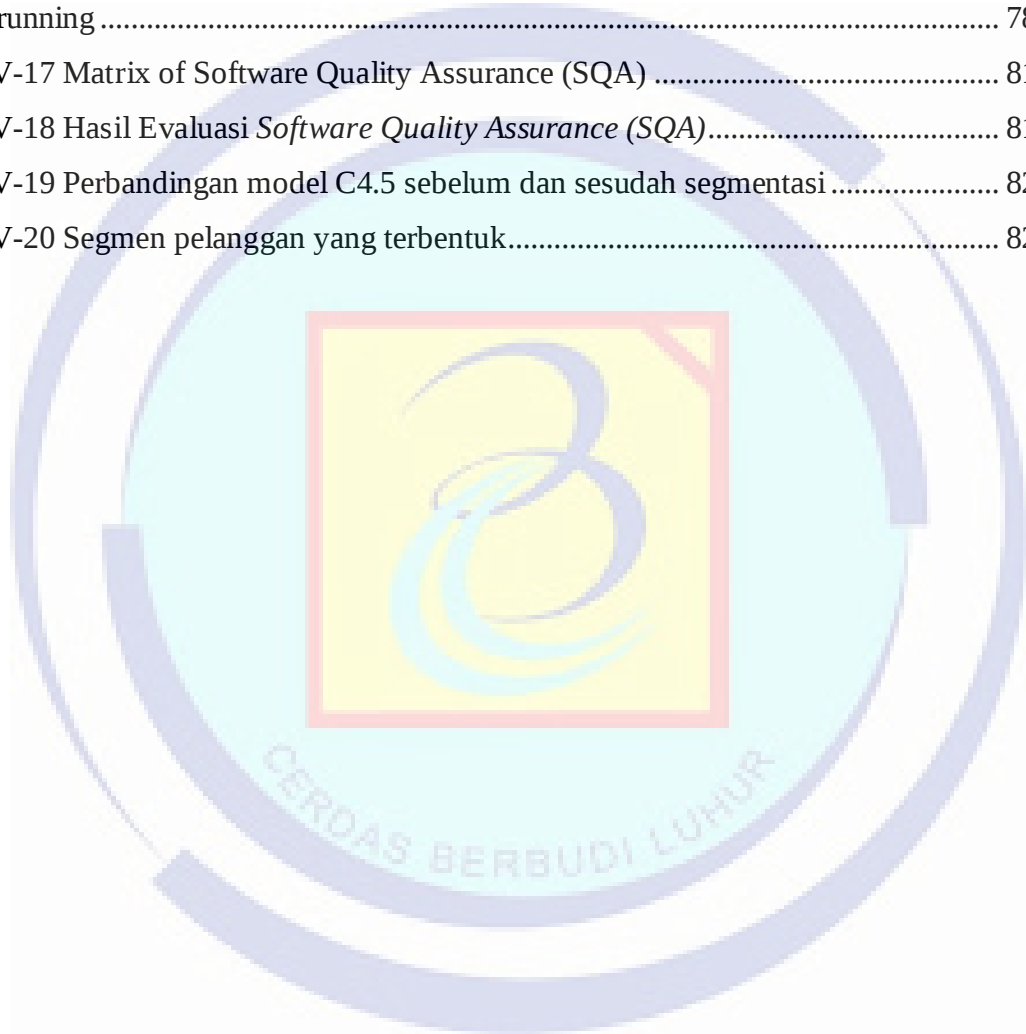
Gambar	Halaman
II-1 Cabang pohon dibuat berdasarkan variasi nilai atribut ^([Han 2006, 295])	11
II-2 Cabang pohon keputusan yang dibuat dari nilai kontinyu ^([Han 2006, 295])	11
II-3 Cabang pohon yang terbentuk dari nilai biner ^([Han 2006, 295])	11
II-4 Pohon keputusan dengan cabang dari atribut A5 tidak konsisten ^([Han 2006], 305)	14
II-5 Pohon keputusan setelah di <i>pruning</i> ^([Han 2006], 305)	15
II-2-6 calon <i>split point</i> untuk <i>temperature</i> ^([Han 2011], 88)	16
II-7 Cabang pohon pertama	17
II-8 Pohon keputusan dengan algoritma C4.5 pada dataset golf ^([Wu 2009], 4)	18
II-9 Pengambilan titik tengah setiap segmen ^([Han 2006], 403)	23
II-10 Pusat segmen bergeser dan objek yang ada dalam segmentasi disesuaikan hingga tidak ada perubahan ^([Han 2006], 403)	23
II-11 Data acak yang akan digunakan sebagai data awal segmen ^([Myatt 2007], 126)	24
II-12 Kumpulan data awal untuk setiap segmen ^([Myatt 2007], 127)	25
II-13 Kerangka pemikiran	34
III-1 Langkah – langkah penelitian	40
III-2 Detail model yang diusulkan	47
III-3 Tampilan pilihan <i>new process</i>	49
III-4 Tampilan <i>import</i> data .csv	50
III-5 Tampilan attribute yang akan diolah	51
III-6 Tampilan anotasi data	52
III-7 Tampilan pemilihan atribut yang akan digunakan	53
III-8 Tampilan operator pemilihan fungsi k - means	53
III-9 Tampilan <i>properties model clustering</i> k - means	54
III-10 Tampilan proses clustering data	54
III-11 Tampilan hasil clustering	54
III-12 Tampilan <i>new building block</i>	55

III-13 Tampilan <i>properties decision tree</i>	56
III-14 Proses penerapan C4.5	56
IV-1 Grafik hasil clustering	65
IV-2 Pohon keputusan prediksi loyalitas pelanggan	69
IV-3 Pohon keputusan yang terbentuk dari cabang <i>disconnect reason(move)</i>	69
IV-4 Pohon keputusan terbentuk dari cabang <i>disconnect reason(not interest)</i>	70
IV-5 Pohon keputusan terbentuk dari cabang <i>disconnect reason(price)</i>	71
IV-6 Pohon keputusan terbentuk dari cabang <i>disconnect reason(service)</i>	72
IV-7 Pohon keputusan terbentuk dari cabang <i>disconnect reason(switch)</i>	72
IV-8 Pohon keputusan terbentuk dari cabang <i>disconnect reason(temporary)</i>	73
IV-9 Pohon keputusan terbentuk dari cabang <i>disconnect reason(trial)</i>	74
IV-10 ROC Curve dari data tanpa attribute potensi pelanggan	78
IV-11 ROC Curve data dengan attribute potensi pelanggan	79
IV-12 Tampilan menu aplikasi prediksi loyalitas pelanggan	80
IV-13 Tampilan proses penginputan pengajuan berhenti pelanggan dari aplikasi prediksi loyalitas pelanggan	80

DAFTAR TABEL

Tabel	Halaman
II-1 Contoh dataset golf ^([Wu 2009], 2)	15
II-2 Nilai informasi dan nilai gain per-atribut pada tahap 1	17
II-3 Nilai informasi untuk semua data yang <i>outlook</i> = <i>sunny</i>	17
II-4 Nilai informasi untuk semua data yang <i>outlook</i> = <i>rain</i>	18
II-5 Data yang akan disegmentasi ^([Myatt 2007], 112)	24
II-6 Nilai titik tengah untuk setiap segmentasi ^([Myatt 2007], 127)	25
II-7 <i>Confusion matrix</i> dengan hasil yang baik.....	26
II-8 Perbandingan penelitian serupa.....	31
III-1 Data awal yang belum di olah.....	37
III-2 Data setelah transformasi	44
III-3 Atribut yang digunakan	44
III-4 Spesifikasi komputer yang digunakan.....	48
III-5 Jadwal penelitian.....	57
IV-1 Contoh data untuk segmentasi pelanggan.....	59
IV-2 Contoh data setelah min-max	60
IV-3 Contoh data sesuai teori RFM	60
VI-4 Hasil perhitungan titik euclidian pertama	61
IV-5 Nilai tengah baru untuk setiap cluster (Tahap 1).....	62
IV-6 Hasil perhitungan titik euclidian kedua.....	62
IV-7 Nilai tengah baru untuk setiap cluster (Tahap II)	63
IV-8 Hasil perhitungan titik euclidian akhir	63
IV-9 Tabel statistic keseluruhan data untuk segmentasi.....	64
IV-10 Nilai titik tengah tiap segmen	64
IV-11 Contoh data untuk klasifikasi C4.5	66
IV-12 Perbandingan hasil akurasi proses dengan pruning dan tanpa pruning	68
IV-13 Perbandingan akurasi <i>confussion matrix</i> dengan pruning dan tanpa pruning per minimal gain	76

IV-14 Hasil <i>confusion matrix</i> dengan rapidminer pada minimal gain 0.06 tanpa attribute hasil <i>k-means</i>	76
IV-15 Hasil <i>confusion matrix</i> dengan rapidminer setelah ditambahkan dengan attribute hasil <i>k-means</i>	77
IV-16 Perbandingan nilai AUC per minimal gain dengan pruning dan tanpa pruning	78
IV-17 Matrix of Software Quality Assurance (SQA)	81
IV-18 Hasil Evaluasi <i>Software Quality Assurance (SQA)</i>	81
IV-19 Perbandingan model C4.5 sebelum dan sesudah segmentasi	82
IV-20 Segmen pelanggan yang terbentuk.....	82



DAFTAR LAMPIRAN

Lampiran	Halaman
1 Dataset	89
2 Angket Penelitian	95



DAFTAR PUSTAKA

- [Afaq 2010] Afaq Alan Khan, Sanjay Jamwal, & M.M. Sepehri . 2009. “Applying Data Mining to Customer Churn Prediction in a Internet Service Provider”. *BGSB University J&K, India*.
- [Aishwarya 2015] Aishwarya Churi, Mayuri Divekar, Sonal Dashpute, & Prajakta Kamble. 2015. “Analysis of Customer Churn in Mobile Industry Using Data Mining”. *International Journal of Emerging Technology and Advanced Engineering*.
- [Ali 2009] Ali S. A, Sulaiman N, Mustapha A, & Mustapha N. 2009. “K-Means Clustering to Improve the Accuracy of Decision Tree Response Classification”. *Information Technology Journal* 8, 8, 1256-1262.
- [Cheng 2008] Cheng, C. H., & Chen, Y. S . 2009. “Classifying The Segmentation of Customer Value Via RFM Model and RS Theory”. *Expert Systems with Applications*, 36, 4176-4184.
- [Dawson 2009] Dawson, C. W. 2009. “Projects in Computing and Information Systems a student’s guide (Second Edition ed.)”. *Harlow, UK: Addison-Wesley*.
- [Gorunescu 2011] Gorunescu, F. 2011. “Data Mining Concepts, Models and Techniques”. Springer-Verlag.
- [Han 2006] Han, J., & Kamber, M. 2006. “Data Mining: Concepts and Techniques (Second Edition ed.)”. *San Francisco: Elsevier Inc.*
- [Han 2011] Han, J., & Kamber, M. 2011. “Data Mining: Practical Machine Learning Tools and Techniques (Third Edition ed.)”. *San Francisco: Elsevier Inc.*
- [John 2006] John Hadden, Ashutosh Tiwari, Rajkumar Roy, & Dymtr Ruta. 2006. “Churn Prediction Using Complaints Data”. *World Academy of Science, Engineering and Technology* 19.
- [Liu 2005] Liu Y, & Schumann, M. 2005. “Data mining feature selection for credit scoring”. *Journal of the Operational Research Society* , 1-10.