

**KAJIAN PENERAPAN ALGORITMA C4.5 UNTUK PENDUGAAN
TINGKAT LOYALITAS PELANGGAN PERUSAHAAN MULTIMEDIA
“XYZ”**

TESIS



Oleh:
Yusuf Kurnia
1111600217

PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR
JAKARTA
2013

YUSUF KURNIA
1111600217

**KAJIAN PENERAPAN ALGORITMA C4.5 UNTUK PENDUGAAN TINGKAT LOYALITAS
PELANGGAN PERUSAHAAN MULTIMEDIA “XYZ”**



TESIS
AGUSTUS
2013

**KAJIAN PENERAPAN ALGORITMA C4.5 UNTUK PENDUGAAN
TINGKAT LOYALITAS PELANGGAN PERUSAHAAN MULTIMEDIA
“XYZ”**

TESIS



Oleh :
Yusuf Kurnia
1111600217

**PROGRAM STUDI : MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCA SARJANA
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2013**

**KAJIAN PENERAPAN ALGORITMA C4.5 UNTUK PENDUGAAN
TINGKAT LOYALITAS PELANGGAN PERUSAHAAN MULTIMEDIA
“XYZ”**

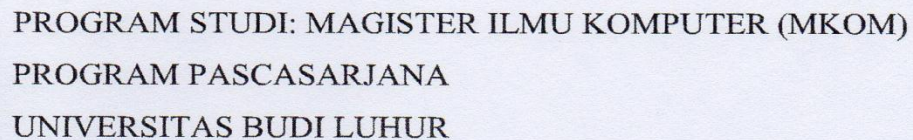
Diajukan untuk memenuhi salah satu persyaratan
memperoleh gelar Magister Ilmu Komputer (MKOM)



Oleh :
Yusuf Kurnia
1111600217

**PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCA SARJANA
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2013**



Saya yang bertanda tangan dibawah ini:

Nama Mahasiswa : Yusuf Kurnia

Nomor Induk Mahasiswa : 1111600217

Program Studi : Magister Ilmu Komputer

Konsentrasi : Teknologi Sistem Informasi

Fakultas/Program : Pascasarjana

Menyatakan bahwa Tesis yang berjudul:

KAJIAN PENERAPAN ALGORITMA C4.5 UNTUK PENDUGAAN TINGKAT LOYALITAS PELANGGAN PERUSAHAAN MULTIMEDIA “XYZ”

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, 27 Agustus 2013





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

LEMBAR PENGESAHAN

Nama Mahasiswa : Yusuf Kurnia
Nomor Induk Mahasiswa : 1111600217
Konsentrasi : Teknologi Sistem Informasi
Jenjang Studi : Magister Ilmu Komputer
Judul Proposal Tesis : KAJIAN PENERAPAN ALGORITMA C4.5 UNTUK
PENDUGAAN TINGKAT LOYALITAS PELANGGAN
PERUSAHAAN MULTIMEDIA "XYZ"

Jakarta, 27 Agustus 2013

Tim Penguji:

Tanda Tangan:

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

Anggota,
(Samidi., M.Kom, M.M)

Pembimbing Utama,
(Dr. Ir. Prabowo Pudjo Widodo., Ms)

Ketua Program Studi

(Dr. Ir. Nazori AZ., MT)

ABSTRAK

Pengguna internet sekarang ini semakin meningkat. Sehingga banyak perusahaan penyedia layanan internet muncul yang mengakibatkan terjadinya persaingan antara perusahaan yang memiliki jumlah pelanggan yang banyak sehingga kecenderungan untuk mempertahankan pelanggan merupakan tantangan besar bagi suatu perusahaan penyedia layanan internet. Untuk dapat mempertahankan pelanggan, maka perlu sebuah model yang dapat melihat loyalitas pelanggan sehingga perusahaan dapat mengenali pelanggan loyal dan tidak loyal dan mampu memprediksi seberapa banyak pelanggan yang akan meninggalkan perusahaan dan berpaling kepada kompetitor. Dengan mengetahui loyalitas pelanggan, perusahaan juga dapat meningkatkan hubungan dengan para pelanggannya. Algoritma C4.5 merupakan algoritma data mining yang cukup baik untuk mengetahui dan memprediksi pola loyalitas pelanggan dengan akurasi yang tinggi. Model C4.5 memiliki akurasi yang sangat tergantung pada atribut pembentuknya, oleh karena itu pengolahan data awal sangat penting untuk dapat menghasilkan model data yang tingkat akurasinya lebih tinggi. Algoritma *Particle Swarm Optimization* (PSO) dan *genetic algorithm* (GA) akan diterapkan untuk pembobotan *attribute* sehingga nilai atribut yang diolah dengan C4.5 dapat disesuaikan untuk mendapatkan hasil yang maksimal. Dengan terbentuknya model loyalitas pelanggan, perusahaan dapat mempertahankan pelanggan baik, dan mampu meningkatkan hubungan terhadap pelanggan yang kurang loyal.

Kata kunci:

Loyalitas pelanggan, pembobotan attribute, algoritma C4.5, algoritma Particle Swarm Optimization (PSO), genetic algorithm (GA)

ABSTRACT

Internet users now been increased. So many internet service providers emerging companies which resulted in a competition between the companies that have a lot of subscribers so that the tendency to retain customers is a major challenge for an internet service provider company. To be able to retain customers, it needs a model that can see customer loyalty so that companies can identify loyal and disloyal customers and be able to predict how many customers are going to leave the company and turned to competitors. By knowing customer loyalty, companies can also improve relationships with customers. C4.5 algorithm is a data mining algorithms are good enough to identify and predict patterns of customer loyalty with high accuracy. C4.5 models have an accuracy that is highly dependent on its constituent attributes, therefore the initial data processing is very important to be able to produce a data model that a higher level of accuracy. Algorithm Particle Swarm Optimization (PSO) and genetic algorithm (GA) will be applied for attribute weighting so that the attribute values are processed by C4.5 can be adjusted to get the maximum results. With the formation of customer loyalty models, companies can retain good customers, and to improve the customer relationships that are less loyal.

Keywords:

Customer loyalty, weighting attributes, C4.5 algorithm, Particle Swarm Optimization algorithm (PSO), genetic algorithm (GA)

LEMBAR PERSEMBAHAN

Satu harta yang sangat berharga yang jarang dimiliki orang masa kini, adalah kejujuran.

(Yusuf Kurnia)

Dengan mengucapkan puji syukur kepada Tuhan Yang Maha Esa, thesis ini kupersembahkan untuk:

- 1 Bapak. Bpk. Abraham (alm) dan Ibu Ermy Samihardja, tercinta yang telah membesarkan aku dan selalu membimbing, mendukung, memotivasi, memberi apa yang terbaik bagiku serta selalu mendoakan aku untuk meraih kesuksesanku.
- 2 Keluarga Besar Abraham terima kasih atas dukungan doa, semangat dan kasih sayang yang senantiasa diberikan kepada penulis.
- 3 Teman – teman dari keluarga besar STMIK Buddhi, 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 hidayahnya kepada penulis, sehingga penulis bisa menyelesaikan penelitian tesis yang berjudul *KAJIAN PENERAPAN ALGORITMA C4.5 UNTUK PENDUGAAN TINGKAT LOYALITAS PELANGGAN PERUSAHAAN MULTIMEDIA “XYZ”*. Tujuan dari penulisan penelitian tesis ini adalah sebagai salah satu syarat untuk menyusun tesis pada Program Studi Magister Ilmu Komputer, Universitas Budi Luhur Jakarta.

Penulis sungguh sangat menyadari, bahwa dalam menyusun penelitian tesis 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 DR. Ir. Prabowo Pudjo Widodo, MS selaku dosen pembimbing tesis yang telah membimbing dan memotivasi penulis dalam mengerjakan proposal penelitian tesis ini.
2. Orang Tua, Bpk. Abraham (alm) dan Ibu Ermy Samihardja terima kasih atas dukungan doa, semangat dan kasih sayang yang senantiasa diberikan kepada penulis.
3. Keluarga Besar Abraham 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 belajar.
5. Teman seperjuangan Aditiya Hermawan dan Andre yang telah berbagi suka dan duka mulai dari awal perkuliahan sampai selesainya proposal ini.

6. Rekan-rekan mahasiswa MKOM Universitas Budi Luhur kelas XA Semester 1 dan 2 dan MKOM Semester 3 konsentrasi Teknologi Sistem Informasi, terima kasih atas kebersamaan, kerja keras dan dukungan semangatnya.
7. Teman – teman dari keluarga besar STMIK Buddhi.
8. Semua keluarga dan teman-teman yang telah memberikan dukungan serta semua pihak yang tidak dapat disebutkan satu persatu.

Penulis menyadari, sebagai makhluk Tuhan yang jauh dari kesempurnaan, bahwa masih banyak kekurangan dari penelitian tesis ini, oleh karena itu penulis mengharapkan kritik dan saran yang bersifat membangun demi kesempurnaan penelitian tesis. Semoga penelitian tesis ini masih dapat memberikan manfaat dari keterbatasannya.

Jakarta, Juli 2013

Yusuf Kurnia

DAFTAR ISI

ABSTRAK	i
<i>ABSTRACT</i>	ii
LEMBAR PERSEMBAHAN	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
DAFTAR LAMPIRAN.....	xiii
BAB I	14
1.1 Latar Belakang	14
1.2 Masalah Penelitian	16
1.2.1 Identifikasi Masalah	16
1.2.2 Pembatasan Masalah	16
1.2.3 Rumusan Masalah	16
1.3 Tujuan dan Manfaat Penelitian.....	17
1.3.1 Tujuan Penelitian	17
1.3.2 Manfaat Penelitian	17
1.4 Tata-Urut Penulisan.....	18
1.5 Daftar Pengertian.....	20
BAB II.....	21
2.1 Tinjauan Pustaka	21
2.1.1 Data Mining	21
2.1.2 Algoritma Klasifikasi	23
2.1.3 Algoritma C4.5.....	25
2.1.4 Algoritma Optimasi.....	36
2.1.5 Algoritma <i>Particle Swarm Optimization</i> (PSO)	37
2.1.6 <i>Genetic Algorithm</i> (GA).....	43
2.1.7 <i>Software Quality Assurance</i> (SQA)	53

2.2	Tinjauan Studi	56
2.2.1	Model Penelitian Daskalaki	56
2.2.2	Model Penelitian Tsai dan Chen	57
2.2.3	Model Penelitian oleh Sousa, Silva dan Neves	59
2.2.4	Rangkuman Penelitian Terkait.....	60
2.3	Tinjauan Obyek Penelitian	61
2.3.1	Segmentasi Pelanggan.....	61
2.3.2	Loyalitas Pelanggan	63
2.3.3	<i>Customer Relationship Management (CRM)</i>	65
2.4	Pola Pikir	67
2.5	Hipotesis.....	69
BAB III		70
3.1	Metodologi Penelitian	70
3.2	Sampling / Metode Pemilihan Data.....	71
3.3	Metode Pengumpulan Data	71
3.3.1	Studi Kepustakaan.....	71
3.3.2	Data Primer	72
3.4	Langkah-Langkah Penelitian.....	72
3.5	Desain Penelitian.....	74
3.5.1	<i>Research Understanding Phase</i>	74
3.5.2	<i>Data Understanding Phase</i>	75
3.5.3	<i>Data Preparation Phase</i>	78
3.5.4	<i>Modelling Phase</i>	79
3.5.5	<i>Evaluation Phase</i>	80
3.5.6	<i>Deployment Phase</i>	88
3.6	Jadwal Penelitian.....	90
BAB IV		92
4.1	Pengelompokan Data, Analisis dan Temuan-temuan.....	92
4.1.1	Proses Penelitian	92
4.1.2	Dataset Perhitungan Manual	92
4.1.3	Perhitungan Algoritma Klasifikasi C4.5.....	95

4.2	Evaluasi dan Validasi	105
4.2.1	Pengukuran Dengan <i>Confusion Matrix</i>	105
4.2.2	Pengukuran AUC (<i>Area Under Curve</i>) dengan ROC (<i>Receiver Operating Characteristic</i>) Curve	113
4.3	Hasil Komparasi	116
4.4	Penerapan Algoritma Terpilih	117
4.5	Implikasi Penelitian	121
4.4.1	Aspek Sistem.....	122
4.4.2	Aspek Manajerial	123
4.4.3	Aspek Penelitian Lanjutan	123
BAB V		125
5.1	Kesimpulan.....	125
5.2	Saran	125
DAFTAR PUSTAKA		126
DAFTAR LAMPIRAN		129
RIWAYAT HIDUP SINGKAT		208

DAFTAR GAMBAR

Gambar	Hal
II-1 Contoh Pohon Keputusan ^([Witten 2011], 99)	26
II-2 Pembuatan Cabang Pohon Keputusan ^([Han 2006], 293)	27
II-3 Pohon Keputusan Yang Dipangkas Dengan <i>Postprunning</i> ^([Alpaydin 2010], 194)	31
II-4 Contoh Repetisi Pohon Keputusan Pada Atribut A1 ^([Alpaydin 2010], 194)	32
II-5 Perulangan Dalam Pohon Keputusan ^([Alpaydin 2010], 194)	32
II-6 Keputusan Setelah Dicabangkan Dengan <i>Age</i>	35
II-7 Pemilihan titik perpecahan dalam atribut	36
II-8 Pergerakan partikel	42
II-9 Roulette Wheel ^([Han 2007], 351)	47
II-10 Bagian-bagian dari SQA (<i>Software Quality Assurance</i>) ^([Pressman 2001], 1)	54
II-11 Model yang diusulkan oleh Daskalaki	57
II-12 Model yang diusulkan oleh Tsai	58
II-13 Model yang diusulkan oleh Sousa	60
II-14 Kerangka pemikiran	68
III-1 Tahap CRISP-DM (<i>Cross Standard Industries Process for Data Mining</i>) ^([Chapman 2000], 1)	73
III-2 Model yang Diusulkan	80
III-3 Membuat Proses Baru	81
III-4 Menentukan Tempat Penyimpanan Data	82
III-5 Menentukan Dataset Yang Ingin Digunakan	82
III-6 Pemberian tipe data <i>attribute</i>	83
III-7 Tampilan <i>data view</i>	84
III-8 Tampilan <i>New Building Block</i>	85
III-9 Relasi <i>Decision Tree</i>	85
III-10 <i>Validation Decision Tree</i> dalam <i>Optimize Weights (PSO)</i>	86
III-11 <i>Decision Tree</i> dengan optimasi <i>Optimize Weights (PSO)</i>	87

III-12 <i>Decision Tree</i> dengan optimasi <i>Optimize Selection (GA)</i>	88
IV-1 Percabangan pertama <i>disc_reason</i>	98
IV-2 Percabangan Kedua Untuk Data Not Interest.....	101
IV-3 Percabangan Kedua Untuk Data Service.....	104
IV-4 Model Loyalitas Pelanggan Dengan Algoritma C4.5.....	107
IV-5 AUC Curve algoritma C4.5.....	114
IV-6 AUC Curve algoritma C4.5 Dengan Optimasi <i>Genetic Algorithm (GA)</i>	115
IV-7 Curve algoritma C4.5 Dengan Optimasi <i>Particle Swarm Optimization (PSO)</i>	115
IV-8 <i>Graphical User Interface (GUI)</i> Sistem Pendugaan Loyalitas Pelanggan.....	119

DAFTAR TABEL

Tabel	Hal
II-1 Data pembelian komputer	34
II-2 Perbandingan penelitian serupa.....	61
III-1 Data awal	76
III-2 <i>Attribute</i> Yang Digunakan	77
III-3 Spesifikasi komputer yang digunakan	81
III-4 Tabel pembobotan <i>attribute</i> SQA (<i>Software Quality Assurance</i>)	89
III-5 Jadwal Penelitian	91
IV-1 Data Yang Digunakan Untuk Contoh Perhitungan C4.5.....	94
IV-2 Perhitungan Info dan Gain Percabangan Pertama	95
IV-3 Perhitungan Entropy dan Gain Percabangan Pertama.....	96
IV-4 Perhitungan Info dan Gain Percabangan Kedua Untuk Data Not Interest	98
IV-5 Perhitungan Entropy dan Gain Percabangan Kedua Untuk Data Not Interest .	100
IV-6 Perhitungan Info dan Gain Percabangan Kedua Untuk Data Service	101
IV-7 Perhitungan Entropy dan Gain Percabangan Kedua Untuk Data Service	103
IV-8 Nilai Accuracy Algoritma C4.5.....	106
IV-9 Nilai Precision Algoritma C4.5	106
IV-10 Nilai <i>Recall</i> Algoritma C4.5.....	107
IV-11 Nilai <i>Accuracy</i> Algoritma C4.5 Dengan Optimasi <i>Genetic Algorithm</i> (GA). 109	
IV-12 Nilai <i>Precision</i> Algoritma C4.5 Dengan Optimasi <i>Genetic Algorithm</i> (GA). 109	
IV-13 Nilai <i>Recall</i> Algoritma C4.5 Dengan Optimasi <i>Genetic Algorithm</i> (GA).....	110
IV-14 Pembobotan <i>Attribute Genetic Algorithm</i> (GA).....	110
IV-15 Nilai <i>Accuracy</i> Algoritma C4.5 Dengan Optimasi <i>Particle Swarm Optimization</i> (PSO).....	111
IV-16 Nilai <i>Precision</i> Algoritma C4.5 Dengan Optimasi <i>Particle Swarm Optimization</i> (PSO).....	111

IV-17 Nilai <i>Recall</i> Algoritma C4.5 Dengan Optimasi <i>Particle Swarm Optimization</i> (PSO).....	112
IV-18 Pembobotan <i>Particle Swarm Optimization</i> (PSO)	112
IV-19 Perbandingan Algoritma C4.5, Algoritma C4.5 + GA dan Algoritma C4.5 + PSO	116
IV-20 Nilai Akurasi Data Baru Algoritma C4.5 yang dioptimasi dengan <i>Particle Swarm Optimization</i> (PSO).....	118
IV-21 Detail Prediksi 20 Data Baru	118
IV-22 <i>Metric of Software Quality Assurance</i> (SQA).....	120
IV-23 Hasil Evaluasi <i>Software Quality Assurance</i> (SQA)	120
IV-24 Spesifikasi <i>Hardware</i> dan <i>Software</i>	122

DAFTAR LAMPIRAN

Lampiran	Hal
1. Data Set Yang Digunakan Dengan 3000 Data	129
2. 10 Angket Pada Penelitian	207

DAFTAR PUSTAKA

- [Alpaydın 2012] Alpaydın, Ethem. *Introduction to Machine Learning*. Second Edition. London: The MIT Press, 2010.
- [Au 2003] Au, Wai Ho, Keith C.C. Chan, and Xin Yao. "A Novel Evolutionary Data Mining Algorithm With Applications To Churn Prediction." *IEEE Transactions on Evolutionary Computation*, 2003: 532-533.
- [Berndtsson 2008] Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. *A Guide For Students In Computer Science And Information Systems*. London: Springer, 2008: 27.
- [Buttle 2007] Buttle, Francis. *Customer Relationship Management* (manajemen hubungan pelanggan) concept and Tools, Jakarta: Bayumedia, 2007: 4-109.
- [Chapman 2000] Chapman, Peter. CRISP-DM 1.0 Step by step Data Mining Guide. SPSS Inc, 2000.
- [Cheng 2009] 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.
- [Daskalaki 2003] Daskalaki, S., Kopanas, I., Goudara, M., & Avouris, N. Data Mining for Decision Support on Customer Insolvency in Telecommunications Business. *European Journal of Operational Research* 145, 2003: 239-255.
- [Dawson 2009] Dawson, C. W. *Projects In Computing And Information System A Student's Guide*. England: Addison-Wesley, 2009: 87.
- [Gorunescu 2011] Gorunescu, Florin. *Data Mining Concepts, Models and Techniques*. Springer-Verlag, 2011.—. *Data Mining: Concepts, Models and Techniques* . Springer, 2011.
- [Han 2007] Han, Jiawei, and Micheline Kamber. *Data Mining: Concepts and Techniques*. Second Edition. San Francisco: Elsevier Inc., 2006.—. *Data Mining: Concepts and Techniques 2nd Edition*. Elsevier, 2007.