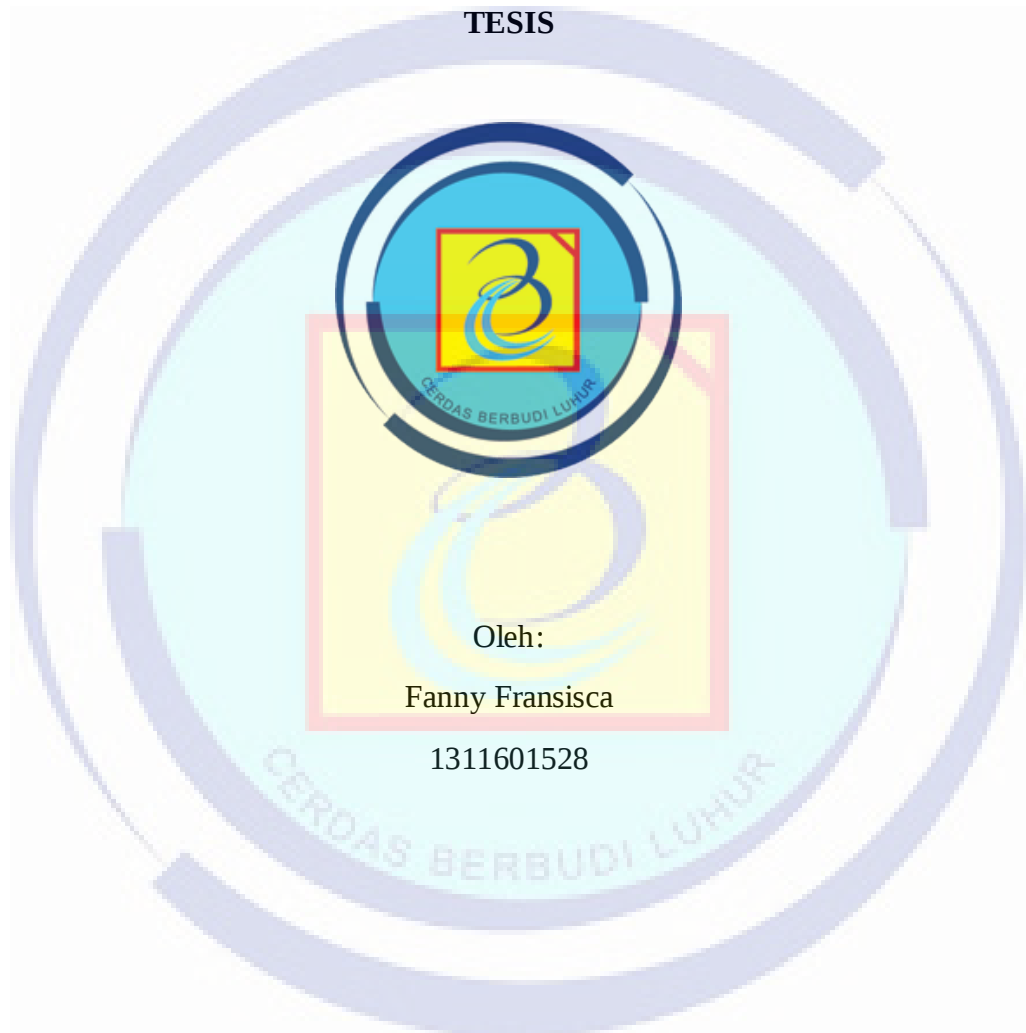


**SEGMENTASI DAN KLASIFIKASI PERILAKU
PEMBAYARAN PELANGGAN PADA PERUSAHAAN
PENYEDIA LAYANAN MULTIMEDIA DENGAN
ALGORITMA K – MEANS DAN C4.5**

TESIS



PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)

PROGRAM PASCASARJANA

UNIVERSITAS BUDI LUHUR

JAKARTA

2016

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

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



Oleh :

Fanny Fransisca

1311601528

PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)

PROGRAM PASCA SARJANA

UNIVERSITAS BUDI LUHUR

JAKARTA

2016

PENGUMUMAN



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

LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama Mahasiswa : Fanny Fransisca
Nomor Induk Mahasiswa : 1311601528
Program Studi : Magister Ilmu Komputer
Konsentrasi : Teknologi Sistem Informasi
Fakultas/Program : Pascasarjana

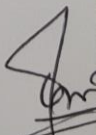
Menyatakan bahwa Tesis yang berjudul:

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



(Fanny Fransisca)



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

LEMBAR PENGESAHAN

Nama Mahasiswa : Fanny Fransisca
Nomor Induk Mahasiswa : 1311601528
Konsentrasi : Teknologi Sistem Informasi
Jenjang Studi : Magister Ilmu Komputer
Judul Proposal Tesis : SEGMENTASI DAN KLASIFIKASI PERILAKU
PEMBAYARAN PELANGGAN PADA
PERUSAHAAN PENYEDIA LAYANAN
MULTIMEDIA DENGAN ALGORITMA K -
MEANS DAN 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

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, datamining, perusahaan jasa*

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, datamining, service company

LEMBAR PERSEMBAHAN

Kepercayaan bagaikan sebuah cermin yang rapuh yang layak untuk di pertahankan.

(Fanny Fransisca)

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

- 1 Bpk. Suhendro dan Ibu Roslinda Usman, 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 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”. Proposal tesis tersebut melengkapi salah satu persyaratan yang diajukan dalam rangka menempuh ujian akhir untuk memperoleh gelar Magister Ilmu Komputer (M.Kom.) pada Program Pasca Sarjana (S2), Program Studi Sistem Informatika di Universitas Budi Luhur.

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 Prof. DR. Moedjiono, M.Sc selaku dosen pembimbing tesis yang telah membimbing dan memotivasi penulis dalam mengerjakan proposal penelitian tesis ini.
2. Kedua orangtua, Bpk. Suhendro dan Ibu Roslinda Usman 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.

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 sistem informasi di Indonesia.

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.....	xii
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	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 Tata Urut Penulisan	5
BAB II LANDASAN TEORI DAN KERANGKA KONSEP	8
2.1 Tinjauan Pustaka.....	8
2.1.1 Data Mining	8
2.1.2 Segmentasi Pelanggan	10
2.1.3 Kelayakan Pelanggan	12
2.1.4 Algoritma Klasifikasi	14
2.1.5 Algoritma Segmentasi.....	26
2.1.6 Evaluasi dan Pengujian Algoritma Klasifikasi	33
2.2 Tinjauan Objek Penelitian.....	34

2.2.1	Prediksi Kelayakan Pelanggan	34
2.2.2	Dataset Perusahaan Layanan Multimedia	35
2.3	Tinjauan Studi.....	35
2.3.1	Model Penelitian Daskalaki ^([Daskalaki 2003], 239)	35
2.3.2	Model Penelitian Pinheiro ^([Pinheiro 2006], 65)	36
2.3.3	Model Penelitian Ali ^([Ali 2009], 1256)	38
2.3.4	Rangkuman Penelitian Terkait	39
2.4	Kerangka Pemikiran	40
2.5	Hipotesis.....	43
BAB III METODE PENELITIAN		44
3.1	Jenis Penelitian	44
3.2	Metode Pengumpulan Data.....	46
3.2.1	Metode Pengolahan Data Awal	46
3.2.2	Studi Kepustakaan.....	46
3.3	Langkah – Langkah Penelitian	46
3.4	Design Penelitian.....	48
3.4.1	Pengumpulan Data.....	48
3.4.2	Pengolahan Data	48
3.4.3	Model/ Metode yang Diusulkan	53
3.4.4	Eksperimen dan Pengujian Model.....	55
3.4.5	Evaluasi dan Validasi Hasil	60
3.5	Jadwal Penelitian	61
BAB IV PEMBAHASAN HASIL PENELITIAN		64
4.1	Hasil Penelitian.....	64
4.1.1	Segmentasi Potensial Pelanggan	64
4.1.2	Klasifikasi Kelayakan Pelanggan	69
4.2	Evaluasi dan Validasi	79
4.2.1	Evaluasi Model Confusion Matrix	80
4.2.2	Kurva ROC(Receiver Operating Characteristic)	82
4.3	Penerapan Algoritma Terpilih	84
4.4	Implikasi Penelitian	87
4.5	Rencana Implementasi	89

BAB V PENUTUP	91
5.1 Kesimpulan	91
5.2 Saran	92
DAFTAR PUSTAKA	93
LAMPIRAN	95
RIWAYAT HIDUP SINGKAT	114



DAFTAR GAMBAR

Gambar	Hal.
II-1 Contoh pohon keputusan ^([Han 2006], 291)	17
II-2 Pembuatan cabang pohon keputusan ^([Han 2006], 295)	18
II-3 Pohon keputusan yang dipangkas dengan postpruning ^([Han 2006], 305)	22
II-4 Contoh repetisi pohon keputusan pada atribut A1 ^([Han 2006], 307)	23
II-5 Perulangan dalam pohon keputusan ^([Han 2006], 307)	23
II-6 Keputusan setelah dicabangkan dengan age ^([Han 2006], 300)	25
II-7 Pemilihan titik perpecahan dalam atribut	26
II-8 Pengambilan titik tengah setiap segmen ^([Han 2006], 403)	30
II-9 Pusat segmen bergeser, dan objek yang ada dalam segmentasi di sesuaikan hingga tidak ada perubahan ^([Han 2006], 403)	31
II-10 Data acak yang akan digunakan sebagai data awal segmen ^([Myatt 2007], 126)	32
II-11 Kumpulan data awal untuk setiap segmen ^([Myatt 2007], 127)	32
II-12 Model yang diusulkan oleh Daskalaki	36
II-13 Model yang diusulkan oleh Pinheir	37
II-14 Model yang diusulkan oleh Ali	39
II-15 Kerangka pemikiran	42
III-1 Langkah – langkah penelitian	48
III-2 Detail model yang diusulkan	54
III-3 Import data dari dokumen CSV	56
III-4 Tampilan atribut dari data yang akan di olah	56
III-5 Parameter untuk segmentasi pelanggan	57
III-6 Relasi antara proses import data dengan proses clustering	58
III-7 Centroid tabel dari cluster model yang terbentuk	58
III-8 Building block untuk C4.5	59
III-9 Parameter untuk proses decision tree	60
III-10 Relasi untuk mengklasifikasikan kelayakan pelanggan	60
IV-1 Pohon keputusan prediksi kelayakan pelanggan	73
IV-2 Pohon keputusan yang terbentuk untuk segmen pertama	73
IV-3 Pohon keputusan untuk segmen kedua	74
IV-4 Pohon keputusan untuk segmen ketiga	75
IV-5 Pohon keputusan untuk segmen keempat	76
IV-6 Pohon keputusan untuk segmen kelima	77
IV-7 ROC Curve untuk dataset sebelum segmentasi	83
IV-8 ROC Curve untuk dataset setelah segmentasi	84
IV-9 Tampilan menu aplikasi prediksi kelayakan pelanggan	85
IV-10 Tampilan view detail data dari tabel awal	85

IV-11 Tampilan proses pemilihan pelanggan untuk di prediksi dengan aplikasi prediksi kelayakan pelanggan.....	86
---	----



DAFTAR TABEL

Tabel	Hal.
II-1 Data pembelian komputer ^([Han 2006], 299)	24
II-2 Data yang akan disegmentasi ^([Myatt 2007], 112)	31
II-3 Nilai titik tengah untuk setiap segmentasi ^([Myatt 2007], 127)	32
II-4 Confusion matrix dengan hasil yang baik	34
II-5 Perbandingan penelitian serupa.....	39
III-1 Data awal yang belum diolah (data untuk rate, balance, pay dan age terdiri dari 6 bulan)	45
III-2 Data setelah transformasi	52
III-3 Atribut yang digunakan	52
III-4 Spesifikasi komputer yang digunakan	55
III-5 Jadwal penelitian.....	61
IV-1 Contoh data untuk segmentasi pelanggan	64
IV-2 Data yang akan disegmentasi dengan k-means	65
IV-3 Hasil perhitungan titik euclidean pertama	66
IV-4 Nilai tengah baru untuk setiap cluster(Tahap 1).....	66
IV-5 Tabel nilai euclidean pada tahap ke 2	67
IV-6 Tabel Statistik Keseluruhan Data Untuk Segmentasi	68
IV-7 Nilai Titik Tengah Setiap Segmen	68
IV-8 Data contoh untuk perhitungan klasifikasi kelayakan pelanggan	70
IV-9 Nilai Pengujian Indikator	72
IV-10 Nilai akurasi pengujian indikator minimal gain dan pruning	81
IV-11 Confusion matrix untuk dataset yang tidak disegmentasi.....	81
IV-12 Confusion matrix untuk data setelah segmentasi	82
IV-13 Nilai AUC pengujian indikator minimal gain dan pruning	83
IV-14 Matrix of Software Quality Assurance (SQA)	86
IV-15 Hasil Evaluasi Software Quality Assurance (SQA).....	87
IV-16 Perbandingan model C4.5 sebelum dan sesudah segmentasi	88

DAFTAR LAMPIRAN

Lampiran	Hal.
1 Dataset.....	94
2 Angket Penelitian	97



DAFTAR PUSTAKA

- [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.
- [Alpaydin 2010] Alpaydın, E. 2010. "Introduction to Machine Learning (Second Edition ed.)". *London: The MIT Press*.
- [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. 2003. "Data Mining for Decision Support on Customer Insolvency in Telecommunications Business". *European Journal of Operational Research*, 145, 239-255.
- [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.*
- [Larose 2005] Larose, Daniel T. 2005. *Discovering Knowledge In Data : An Introduction to Data Mining*. Canada: John Wiley & Sons, Inc., Hoboken, New Jersey.
- [Liao 2007] Liao, T. W., & Triantaphyllou, E. 2007. "Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications (Vol. 6)". 5 Toh Tuck Link: *World Scientific Publishing Co. Pte. Ltd.*