

**KAJIAN KOMPARASI PENERAPAN ALGORITMA C4.5, *NEURAL NETWORK* DAN *K-NEAREST NEIGHBOR* DALAM PENENTUAN
KELAYAKAN KREDIT KENDARAAN BERMOTOR :
Studi Kasus PT. RAZ**

TESIS



Oleh :

RENNY AISYAH

1111600662

PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER (MKOM)

UNIVERSITAS BUDI LUHUR

JAKARTA

2013

**KAJIAN KOMPARASI PENERAPAN ALGORITMA C4.5, *NEURAL NETWORK* DAN *K-NEAREST NEIGHBOR* DALAM PENENTUAN
KELAYAKAN KREDIT KENDARAAN BERMOTOR :
Studi Kasus PT. RAZ**

TESIS

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



Oleh :

RENNY AISYAH

1111600662

PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER (MKOM)

UNIVERSITAS BUDI LUHUR

JAKARTA

2013



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

LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama Mahasiswa : RENNY AISYAH
Nomor Induk Mahasiswa : 1111600662
Program Studi : Magister Ilmu Komputer
Konsentrasi : Teknologi Sistem Informasi

Menyatakan bahwa TESIS yang berjudul

KAJIAN KOMPARASI PENERAPAN ALGORITMA C4.5, NEURAL NETWORK DAN K-NEAREST NEIGHBOR DALAM PENENTUAN KELAYAKAN KREDIT KENDARAAN BERMOTOR : Studi Kasus PT. RAZ

1. Merupakan hasil karya tulis ilmiah sendiri dan bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik 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 kemudian hari pernyataan ini tidak benar.

Jakarta, 27 Agustus 2013



Renny Aisyah



**PROGRAM STUDI MAGISTER ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

LEMBAR PENGESAHAN

Nama : RENNY AISYAH
Nomor Induk Mahasiwa : 1111600662
Konsentrasi : Teknologi Sistem Informasi
Jenjang Studi : Strata 2
Judul : Kajian Komparasi Penerapan Algoritma C4.5, *Neural Network* dan *K-Nearest Neighbor* Dalam Penentuan Kelayakan Kredit Kendaraan Bermotor : Studi Kasus PT. RAZ

Jakarta, 27 Agustus 2012

Tim Penguji:

Tanda tangan:

Ketua
Dr. Ir. Nazori, AZ, MT

Anggota
Samidi, MM, M.Kom

Pembimbing,
Dr. Ir. Prabowo Pudjowidodo, MS

Ketua Program Studi

Dr. Ir. Nazori, AZ, MT

ABSTRAK

Sejalan dengan pertumbuhan dan perkembangan bisnis, masalah perkreditan tetap menarik untuk dikaji dan diungkap. Beberapa penelitian bidang komputer telah banyak dilakukan untuk mengurangi resiko kredit yang menyebabkan kerugian pada perusahaan. Dalam penelitian ini dilakukan komparasi algoritma C4.5, *Neural Network* dan *K-Nearest Neighbor* yang diaplikasikan terhadap data konsumen yang mendapat pembiayaan kredit motor baik yang bermasalah dalam pembayaran angsurannya maupun tidak. Dari hasil pengujian dengan mengukur kinerja ketiga algoritma tersebut menggunakan metode pengujian *Cross Validation*, *Confusion Matrix* dan Kurva ROC, diketahui bahwa algoritma C4.5 memiliki nilai *accuracy* paling tinggi, yaitu 83.22%, diikuti oleh metode *K-Nearest Neighbor* dengan *accuracy* sebesar 80.74% dan yang terendah adalah metode *neural network* dengan nilai *accuracy* 75.74%. Nilai AUC untuk metode C4.5 juga menunjukkan nilai tertinggi, yaitu 0.956 disusul metode C4.5 dengan nilai AUC sebesar 0.956, dan yang terendah adalah nilai AUC *neural network* yaitu 0.719. Melihat nilai AUC dari ketiga metode tersebut maka ketiganya termasuk kelompok klasifikasi sangat baik karena nilai AUC-nya antara 0.90-1.00.

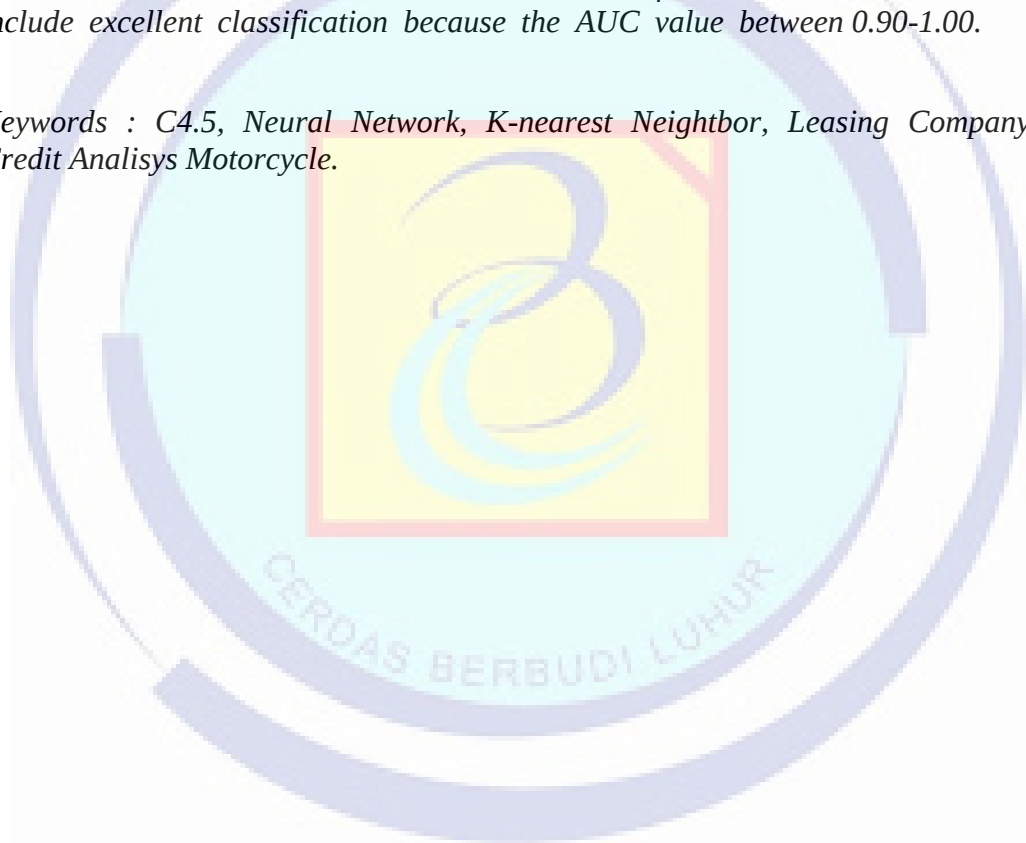
Kata kunci :

C4.5, *Neural Network*, *K-nearest Neighbor*, Perusahaan *Leasing* dan Analisa Kredit Kendaraan Bermotor.

ABSTRACT

In line with the growth and business development, credit issues remain to be studied and revealed interesting. Some of the research field of computers has done much to reduce the credit risk of causing hsrn to the company. In this study a comparison algorithm C4.5, Neural Network dan K-Nearest Neighbor are applied to the data of consumers who have good credit financing motorcycle that consumers are troubled or not. From the test results to measure the performance of the three algorithms using the test method Cross Validation, Confusion Matrix and ROC curves, it is known that the algorithm C4.5 has the highest accuracy value of 83.22%, followed by K-Nearest Neighbor method 80.74%, and the lowest is the neural network method 75.74% . AUC values for C4.5 method also showed the highest value, namely 0.956 and the lowest is the neural network method with AUC values of 0.719. All methodes are include excellent classification because the AUC value between 0.90-1.00.

Keywords : C4.5, Neural Network, K-nearest Neightbor, Leasing Company, Credit Analisis Motorcycle.



KATA PENGANTAR

Dengan memanjatkan Puji dan syukur kehadiran Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya kepada penulis, sehingga penulis bias menyelesaikan proposal tesis ini yang berjudul “Kajian Komparasi Penerapan Algoritma C4.5, *Neural Network* dan *K-Nearest Neighbor* dalam penentuan kelayakan kredit kendaraan bermotor. Tujuan dari penulisan tesis ini untuk memenuhi salah satu syarat dalam menyelesaikan jenjang Strata Dua (S-2) pada Program Studi Magister Ilmu Komputer (MKOM) Program Pascasarjana Universitas Budi Luhur.

Terselesaikannya Tugas Akhir ini dengan baik berkat dukungan, motivasi, petunjuk dan bimbingan dari berbagai pihak. Maka, dalam kesempatan ini penulis mengucapkan terima kasih kepada :

1. Bapak Dr. Moedjiono, M.Sc., sebagai Direktur Program Pascasarjana Universitas Budi Luhur.
2. Bapak Dr. Ir. Nazori AZ, MT. sebagai Ketua Program Studi Magister Ilmu Komputer Universitas Budi Luhur
3. Bapak Prof. DR. Ir. Prabowo, MS selaku dosen pembimbing tesis yang telah membimbing, memotivasi serta memberi waktu dan arahan kepada penulis selama penulisan tesis ini.
4. Yang tercinta kedua orang tua saya Bpk. Patia Raja SH, M.pd dan Ibu Husnawati dan saudara–saudara perempuan saya, Azizah, Ria, Yuli dan Diana beserta seluruh keluarga besar saya yang telah memberikan dukungan moral, spiritual, materi serta doa bagi penulis sehingga penulisan tesis ini dapat terselesaikan.
5. Yang tersayang Zul Ashri yang selalu setia membantu dan memberikan semangat, dukungan, kasih sayang dan doa bagi penulis.
6. Semua Bapak dan Ibu Dosen beserta staf Universitas Budi Luhur di Program Studi Magister Ilmu Komputer Universitas Budi Luhur yang telah mengajarkan dan memberikan banyak ilmu pengetahuan, pencerahan dan bimbingan dalam belajar.

7. Untuk Pak Lukman yang telah membantu penulis dalam menyelesaikan tesis ini.
8. Rekan–rekan mahasiswa MKOM Universitas Budi Luhur XA dan XB teman seperjuangan yang telah menyumbangkan ide dan sarannya, terima kasih atas kebersamaan, kerja keras dan dukungan semangatnya.
9. Rekan–rekan BRI Unit Pondok Pinang yang telah memberikan waktu untuk bimbingan tesis, terima kasih atas kebersamaan, kerja keras dan dukungan semangatnya.
10. Semua pihak yang tak bisa disebutkan satu persatu, yang telah menjadi motivasi demi terselesaikannya penyusunan tesis ini.

Penulis menyadari bahwa tidak ada sesuatu pun yang sempurna kecuali Allah SWT. Oleh karena itu, dengan senang hati penulis menerima kritik dan saran yang bersifat membangun demi kesempurnaan penelitian tesis nantinya. Semoga tesis ini bermanfaat bagi penulis khususnya juga bagi pembaca.

Jakarta, Agustus 2013

Renny Aisyah

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR.....	ii
DAFTAR ISI.....	iv
DAFTAR GAMBAR.....	vi
DAFTAR TABEL	viii
 BAB I : PENDAHULUAN	
1.1 Latar Belakang Penulisan	1
1.2 Identifikasi, Batasan dan Rumusan Masalah	2
1.2.1 Identifikasi Masalah	2
1.2.2 Batasan Masalah.....	2
1.2.3 Rumusan Masalah	3
1.3 Tujuan dan Manfaat Penelitian	3
1.3.1 Tujuan Penelitian	3
1.3.2 Manfaat Penelitian	3
1.4 Sistematika Penulisan	4
 BAB II : LANDASAN TEORI DAN KERANGKA KONSEP	
2.1 Tinjauan Pustaka	5
2.1.1 <i>Data Mining</i>	5
2.2.2 Algoritma Klasifikasi <i>Data Mining</i>	10
2.1.3 Algoritma C4.5.....	15
2.1.3.1 Konsep Algoritma C4.5	15
2.1.3.2 Studi Kasus C4.5.....	16
2.1.4 <i>Neural Network</i>	19
2.1.4.1 Studi Kasus <i>Neural Network</i>	23
2.1.5 <i>K-Nearest Neighbor</i>	25
2.1.6 <i>Rapid Miner 5</i>	32
2.1.7 Evaluasi dan Validasi Metode Klasifikasi <i>Data Mining</i>	32
2.2 Tinjauan Studi	34

2.3	Tinjauan Obyek Penelitian	36
2.3.1	Kredit Pemilikan Kendaraan	36
2.3.1.1	<i>Leasing</i>	36
2.3.1.2	Kredit	36
2.3.1.3	Analisis Kredit	37
2.4	Kerangka Pemikiran.....	38
2.5	Hipotesis Penelitian.....	41

BAB III : METODOLOGI PENELITIAN

3.1.	Jenis Penelitian.....	42
3.2.	Metode Pemilihan Sampel	43
3.2.1.	Populasi.....	43
3.2.2.	Sampel.....	43
3.3.	Metode Pengumpulan Data	44
3.4.	Teknik Analisis Data.....	44
3.4.1.	<i>Data Understanding</i>	45
3.4.2.	Instrument Penelitian	45
3.4.3.	<i>Modeling</i>	47
3.4.4.	<i>Evaluation</i>	47
3.4.5.	<i>Deployment</i>	47
3.4.6.	Jadwal Penelitian	48
3.5	Jadwal Penelitian.....	51

BAB IV : PEMBAHASAN HASIL PENELITIAN

4.1.	Pengelompokan Data	53
4.2	Perancangan Sistem, Prototipe Model, Perencanaan Strategi dan Pengujian Sistem	55
4.2.1	Model C4.5	55
4.3	Implikasi Penelitian.....	74
4.3.1	Aspek Sistem.....	74
4.3.2	Aspek Manajerial	75
4.3.3	Aspek Penelitian Lanjut	75

4.3.4 Rencana Implementasi	75
4.2.2 Algoritma C4.5	57
4.2.3 Model <i>K-Nearest Neighbor</i>	60
4.2.3.1 Penerapan Algoritma <i>K-Nearest Neighbor</i>	64
4.2.4 Evaluasi dan Validasi	66
4.2.4.1 Pengujian Model	66
4.2.5 Penerapan Algoritma Terpilih.....	72
 BAB V : PENUTUP	
5.1 Kesimpulan.....	78
5.2 Saran.....	78
DAFTAR PUSTAKA.....	80
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2.1	Proses KDD	7
Gambar 2.2	Langkah-langkah dalam Proses Data Mining	7
Gambar 2.3	Pohon Keputusan C4.5 dengan Simpul Akar Asset	19
Gambar 2.4	MLP untuk Credit Risk	23
Gambar 2.5	Rapid Miner 5 interface.....	32
Gambar 2.6	Kerangka Berpikir.....	42
Gambar 3.1	Langkah-langkah penelitian.....	50
Gambar 4.1.	Desain model neural network	48
Gambar 4.2	Hasil Neural net	48
Gambar 4.3.	Desain model Validasi	67
Gambar 4.4a	Kurva ROC Algoritma C4.5	69
Gambar 4.4b	Kurva ROC Algoritma neural net.....	70
Gambar 4.4c	Kurva ROC Algoritma K-Nearest Neighbor	70
Gambar 4.5	Interface Beasiswa	73

DAFTAR TABEL

Tabel 2.1	Tabel Data untuk Klasifikasi Credit Risk	16
Tabel 2.2	Nilai Entropy dan Gain untuk menentukan Simpul Akar.....	18
Tabel 2.3	Nilai Awal untuk Input, Bobot dan Bias	23
Tabel 2.4	Perhitungan untuk bobot dan bias terbaru	25
Tabel 2.5	Tabel Kasus.....	26
Tabel 2.6	Definisi Bobot Atribut	27
Tabel 2.7	Kedekatan Nilai Atribut Jenis Kelamin	27
Tabel 2.8	Kedekatan Nilai Atribut Pendidikan.....	27
Tabel 2.9	Kedekatan Nilai Atribut Pendidikan.....	27
Tabel 2.10	Kedekatan Nilai Atribut Status	28
Tabel 2.11	Model Confusion Matrix	33
Tabel 3.1.	Spesifikasi Perangkat keras dan Perangkat Lunak	42
Tabel 3.2	Atribut dan nilai kategori	45
Tabel 3.3.	Sampel Data Training	47
Tabel 3.4.	Metric of Software Quality Assurance (SQA)	48
Table 3.5	Jadwal Penelitian	51
Tabel 4.1	Atribut dan nilai kategori.....	53
Tabel 4.2	Hasil nilai entropy dan gain untuk menentukan simpul akar dengan data training.....	56
Tabel 4.3	Nilai Bobot Akhir untuk Output Layer.....	57
Tabel 4.4.	Defisini bobot atribut pada data training	60
Tabel 4.5	Kedekatan nilai atribut Status perkawinan	60
Tabel 4.6	Kedekatan nilai atribut jumlah tanggungan.....	60
Tabel 4.7	Kedekatan nilai atribut pendidikan terakhir.....	61
Tabel 4.8	Kedekatan nilai atribut Usia	62
Tabel 4.9	Kedekatan nilai atribut kepemilikan rumah.....	62
Tabel 4.10	Kedekatan nilai atribut lama tinggal.....	63
Tabel 4.11	Kedekatan nilai atribut kondisi rumah	63
Tabel 4.12	Kedekatan nilai atribut status kepegawaian.....	63
Tabel 4.13	Kedekatan nilai atribut masa kerja	64

Tabel 4.14	Kedekatan nilai atribut penghasilan perbulan	64
Tabel 4.15	perbandingan nilai data testing 01 dan data training	65
Tabel 4.16	Model Confusion Matrik Metode C4.5	67
Tabel 4.17	Model Confusion Matrik Metode Neural Network	68
Tabel 4.18	Model Confusion Matrik Metode K-Nearest Neighbor	68
Tabel 4.19	Komparasi Nilai Accuracy, Precision, dan Recall	69
Tabel 4.20	Komparasi Nilai AUC	71
Tabel 4.21	Komparasi Nilai Accuracy dan AUC	71
Tabel 4.22	Data Baru untuk Penerapan Algoritma Terpilih	72
Tabel 4.23	Confussion Matrix data baru dengan algoritma C 4.5	72
Tabel 4.24	Rencana Implementasi Sistem Kelayakan Pemberian Kendaraan Bermotor	76
Tabel 4.25	Hardware dan Software	77

DAFTAR LAMPIRAN

LAMPIRAN 1 DATA KONSUMEN KREDIT KENDARAAN BERMOTOR

LAMPIRAN 2 HASIL ALGORITMA C45 DI RAPIDMINER

LAMPIRAN 3 HASIL ALGORITMA NEURAL NET (MLP)

LAMPIRAN 4 K-NN

LAMPIRAN 5 DAFTAR RIWAYAT HIDUP



DAFTAR PUSTAKA

- [Alpaydin2010] Alpaydin, Ethem. (2010). *Introduction to Machine Learning*. London: The MIT Press
- [Angelini 2008] Angelini, Eliana. (2008). *A Neural Network Approach for Credit Risk Evaluation..The Quarterly Review of Economics and Finance*
- [Arminger 1997] Arminger, Gerhard. (1997). *Analyzing CreditRisk Data: A Comparison of Logistic Discrimination, Classification Tree Analysis, and Feedforward Networks* COMPUTATIONAL STATISTICS, Vol 12 No. 2, March 26, 1997
- [Baesens 2001] Baesens, Bart. (2001). *Building Credit Risk Evaluation Expert System Using Neural Network Rule Extraction and Decision Tables*. Management Science, Volume 49, Number 3, pp. 312-329, March 2003.
- [Bramer 2007] Bramer, Max. (2007). *Principles of Data Mining*. London: Springer
- [Gorunescu 2011] Gorunescu, Florin (2011). *Data Mining: Concepts, Models, and Techniques* Verlag Berlin Heidelberg: Springer
- [Han 2006] Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.
- [Heiat 2011] Heiat, Abbas. (2011). *Modeling Consumer Credit Scoring Through Bayes Network*. World Journal of Social Sciences, Volume 1, Number 3, pp.132-141 July 2011
- [Kusrini 2009] Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- [Larosem2005] Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- [Lai 2006] Lai, Kin Keung. (2006). *Credit Risk Analysis Using a Reliability-Based Neural Network Ensemble Model*. ICANN
- [Liao 2007] Liao. 2007). *Recent Advances in Data Mining of Enterprise Data: Algorithms and Application*. Singapore: World Scientific Publishing
- [Maimon2005] Maimon, Oded & Rokach, Lior. (2005). *Data Mining and Knowledge Discovey Handbook*. New York: Springer
- [Myatt2007] Myatt, Glenn J. (2007). *Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*. New Jersey: John Wiley & Sons, Inc.