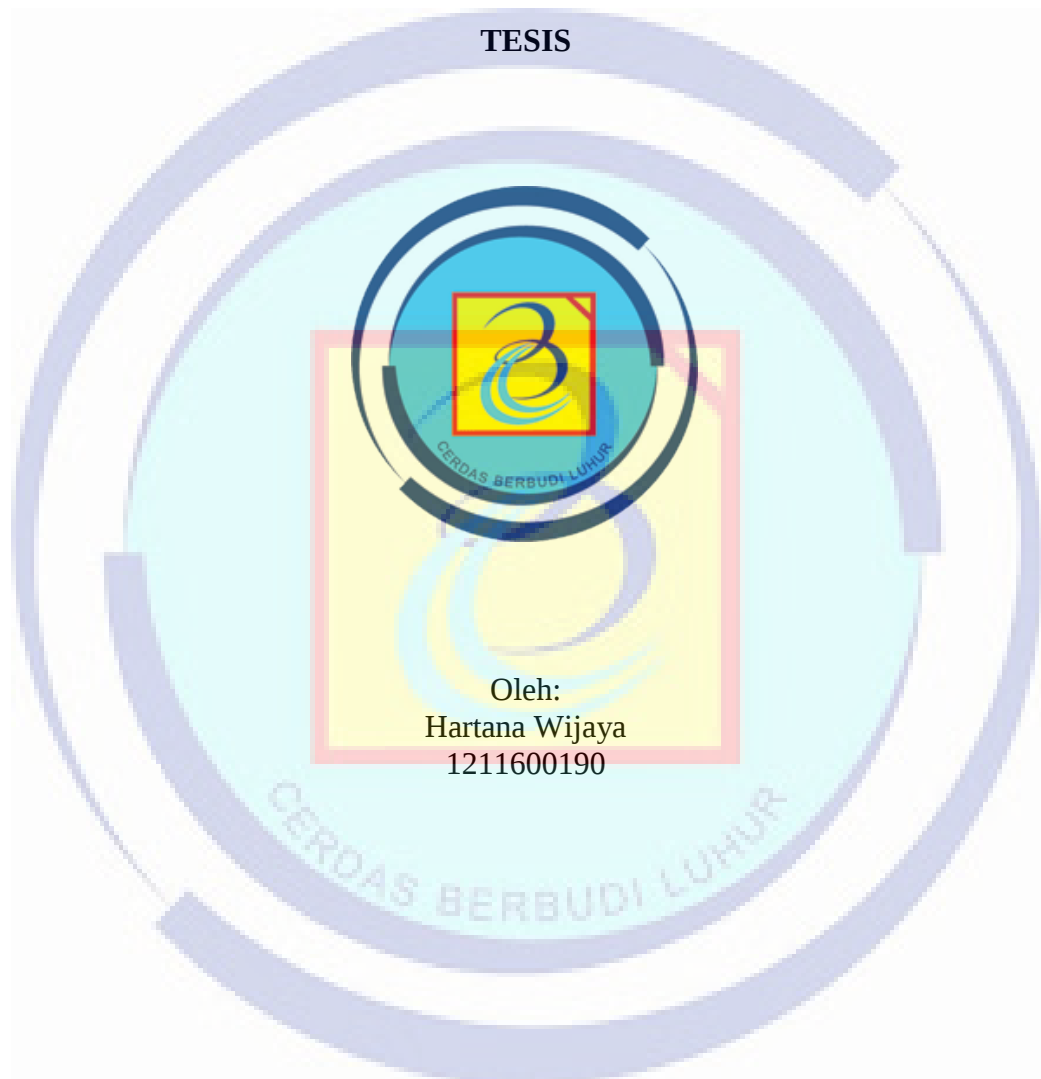


**OPTIMASI PENERAPAN ALGORITMA GENETIKA DENGAN
MENGUNAKAN METODE C4.5 UNTUK MEMPREDIKSI
PENYAKIT KANKER PAYUDARA**



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

**JAKARTA
2014**

**OPTIMASI PENERAPAN ALGORITMA GENETIKA DENGAN
MENGUNAKAN METODE C4.5 UNTUK MEMPREDIKSI
PENYAKIT KANKER PAYUDARA**

TESIS

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



Oleh:
Hartana Wijaya
1211600190

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

**JAKARTA
2014**



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

Lembar Pernyataan

Saya yang bertandatangan dibawah ini:

Nama Mahasiswa : Hartana Wijaya
Nomor Induk Mahasiswa : 1211600190
Program Studi : Magister Ilmu Komputer
Konsentrasi : Teknologi Sistem Informasi
Fakultas/Program : Pascasarjana

menyatakan bahwa Tesis yang berjudul:

***OPTIMASI PENERAPAN ALGORITMA GENETIKA DENGAN
MENGUNAKAN METODE C4.5 UNTUK MEMPREDIKSI PENYAKIT
KANKER PAYUDARA.***

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

Pernyataan ini saya buat dengan penuh tanggungjawab dan saya bersedia menerima konsekuensi apapun sesuai aturan yang berlaku apabila dikemudian hari pernyataan ini tidak benar.

Jakarta 19 Agustus 2014


METRAI
TEMPEL
C11F2ACF106884940
ENAM RIBU RUPAH
6000
DJP
(Hartana Wijaya)



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

Lembar Pengesahan

Nama Mahasiswa : Hartana Wijaya
Nomor Induk Mahasiswa : 1211600190
Konsentrasi : Teknologi Sistem Informasi
Judul Proposal Tesis : OPTIMASI PENERAPAN ALGORITMA
GENETIKA DENGAN MENGGUNAKAN
METODE C4.5 UNTUK MEMPREDIKSI
PENYAKIT KANKER PAYUDARA.

Jakarta, 19 Agustus 2014

Tim Penguji:

Tandatangan:

Ketua,
(Dr., Ir., Nazori Az, M.T)

Anggota,
(Mardi Hardjianto, M.Kom)

Pembimbing Utama,
(Dr. Moedjiono, M.Sc)

Ketua Program Studi Magister Ilmu Komputer,

(Dr., Ir., Nazori Az, M.T)

ABSTRAK

Kanker merupakan tantangan besar bagi umat manusia. Kanker dapat mempengaruhi berbagai bagian tubuh. Penyakit mematikan ini dapat ditemukan pada manusia dari segala usia. Namun, risiko kanker meningkat dengan peningkatan usia. Kanker payudara merupakan penyakit kanker paling umum di kalangan wanita, dan merupakan penyebab kematian terbesar juga bagi wanita. Kemudian terjadi masalah dalam pendeteksian kanker payudara, sehingga menyebabkan pasien mengalami pengobatan yang tidak perlu dan biaya yang besar. Pada penelitian yang serupa, ada beberapa metode yang digunakan tetapi terdapat masalah dikarenakan bentuk dari sel-sel kanker yang nonlinier. Metode C4.5 dapat mengatasi permasalahan tersebut, tetapi C4.5 lemah dalam hal penentuan nilai parameter, sehingga perlu dioptimasi. Algoritma Genetika adalah salah satu metode optimasi yang baik, oleh karena itu nilai-nilai parameter dari C4.5 akan dioptimasi dengan menggunakan Algoritma Genetika sehingga mendapatkan nilai parameter yang paling baik. Hasil dari penelitian ini adalah bahwa Algoritma C4.5 berbasis optimasi algoritma genetika memiliki nilai akurasi yang lebih tinggi (96%) dibanding hanya dengan menggunakan algoritma C4.5 (94.99%) dan yang dioptimasi dengan algoritma PSO (95.71%). Hal ini terbukti dari peningkatan nilai akurasi sebesar 1,01% untuk model algoritma C4.5 yang telah dioptimasi dengan algoritma genetika. Sehingga dapat disimpulkan bahwa penerapan teknik optimasi algoritma genetika dapat meningkatkan nilai akurasi pada algoritma C4.5.

Kata Kunci: Kanker Payudara, Optimasi, Akurasi, C4.5, Algoritma Genetika, Prediksi.

ABSTRACT

Cancer is a major challenge for mankind. Cancer can affect various parts of the body. This deadly disease can be found in humans of all ages. However, the risk of cancer increases with increasing age. Breast cancer is the most common cancer among women, and is also the biggest cause of death for women. Then there are problems in the detection of breast cancer, thus causing the patient to experience unnecessary treatment and a big expense. In similar studies, there are several methods that are used but there is a problem due to the shape of cancer cells are nonlinear. C4.5 method can overcome these problems but C4.5 weak in terms of determining the value of the parameter, so it needs to be optimized. Genetic Algorithm optimization method is one of the good, therefore the values of the parameters of C4.5 will be optimized using Genetic Algorithm thus obtain the best parameter values. Results from this study is that the C4.5 algorithm optimization based genetic algorithm has a higher accuracy rate (96%) compared to only using the C4.5 algorithm (94.99%) and optimized with PSO algorithm (95.71%). This is evident from the increase in value by 1.01% accuracy for the C4.5 algorithm model has been optimized with a genetic algorithm. It can be concluded that the application of genetic algorithm optimization techniques can improve the accuracy of the algorithm C4.5.

Keywords: Breast Cancer, Optimization, accuracy, C4.5, Genetic Algorithms, Predictions.

LEMBAR PERSEMBAHAN

*Tidak melakukan kejahatan, tambah perbuatan baik, sucikan hati dan pikiran.
(Buddha)*

Dengan mengucapkan puji dan syukur kehadiran Tuhan Yang Maha Esa dan Sang Triratna, tesis ini saya persembahkan dengan penuh suka cita untuk:

1. Papa dan Mama tercinta, yang sudah membesarkan, merawat dan mendidik dengan penuh cinta dan kasih sayang, serta senantiasa memberikan nasihat, arahan, dukungan dan doa demi untuk kesuksesanku.
2. Kakak dan adikku, yang selalu senantiasa memberikan dukungan dan doanya.
3. Istri tercinta, yang sudah senantiasa sabar dan menemaniku disaat suka dan duka.
4. Anak-anakku tersayang, semoga kalian senantiasa akan menjadi anak yang baik, berbakti kepada kedua orangtua serta menjadi anak yang berguna bagi orang lain di kemudian hari.
5. Teman-teman semua di STMIK Buddhi, Tangerang, yang senantiasa memberikan doa dan dukungannya.
6. Semua pihak yang tidak dapat saya sebutkan satu persatu, yang telah memberikan dukungannya.

KATA PENGANTAR

Dengan memanjatkan Puji dan syukur kehadiran Tuhan Yang Maha Esa dan Sang Triratna yang telah melimpahkan segala rahmat dan berkahnya kepada penulis, sehingga penulis bisa menyelesaikan penelitian yang berjudul *Optimasi Penerapan Algoritma Genetika Dengan Menggunakan Metode C4.5 Untuk Memprediksi Penyakit Kanker Payudara*. Tujuan dari penulisan penelitian 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 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 DR. Moedjiono, M.Sc selaku dosen pembimbing tesis yang telah membimbing dan memotivasi penulis dalam mengerjakan proposal penelitian tesis ini.
2. Kedua orangtua, Bpk. Asan Widjaja dan Ibu Titih terima kasih atas dukungan doa, semangat dan kasih sayang yang senantiasa diberikan kepada penulis.
3. Keluarga Besar Oey terima kasih atas dukungan doa, semangat dan kasih sayang yang senantiasa diberikan kepada penulis.
4. Istri dan anak-anak tercinta yang selalu memberikan semangat, dukungan, kasih sayang dan doa bagi penulis.
5. 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.
6. Rekan-rekan mahasiswa MKOM Universitas Budi Luhur angkatan 2012.
7. Teman-teman dari keluarga besar STMIK Buddhi 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.

Jakarta, Agustus 2014

Hartana Wijaya



DAFTAR ISI

	Halaman
ABSTRAK	i
ABSTRACT	ii
LEMBAR PERSEMBAHAN	iii
KATA PENGANTAR	iv
DAFTAR ISI	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	2
1.2.1 Identifikasi Masalah	2
1.2.2 Pembatasan Masalah	3
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 Tata-Urut Penulisan	4
1.5 Daftar Pengertian	5
BAB II LANDASAN TEORI DAN KERANGKA KONSEP	7
2.1 Tinjauan Pustaka	7
2.1.1 Data Mining	7
2.1.2 Algoritma Klasifikasi Data Mining	9
2.1.3 Algoritma Optimasi	13
2.1.4 Algoritma C4.5	13
2.1.5 Algoritma <i>Particle Swarm Optimization</i> (PSO)	19
2.1.6 Algoritma Genetika	23
2.1.7 Pemilihan Variabel	30

2.1.8	Evaluasi dan Validasi Metode Klasifikasi Data Mining	31
2.1.9	Kanker Payudara	33
2.1.10	<i>Software Quality Assurance</i>	36
2.2	Tinjauan Studi	38
2.2.1	Penelitian oleh Yaoying, et.al ^[Yaoying 2011]	38
2.2.2	Penelitian oleh Afzan Adam dan Khairuddin Omar ^[Adam 2006]	40
2.2.3	Penelitian oleh M. Lundin, et.al ^[Lundin 1999]	42
2.2.4	Penelitian oleh Abdelghani Bellachia dan Erhan Guven ^[Bellachia 2000]	43
2.2.5	Rangkuman Penelitian Terkait	44
2.3	Tinjauan Obyek Penelitian	45
2.3.1	Prediksi Kanker Payudara	45
2.3.2	<i>UCI Machine Learning Repository Dataset</i>	45
2.4	Pola Pikir	46
2.4.1	Kerangka Konsep	46
2.4.2	Kerangka Pemikiran Pemecahan Masalah	47
2.5	Hipotesis	47
BAB III METODOLOGI DAN DESAIN PENELITIAN		49
3.1	Jenis Penelitian	49
3.2	Metode Pengumpulan Data	52
3.2.1	Data Sekunder	52
3.2.2	Metode Pengolahan Awal Data	52
3.2.3	Studi Kepustakaan	52
3.3	Langkah-Langkah Penelitian	52
3.4	Desain Penelitian	53
3.4.1	Pengumpulan Data	53
3.4.2	Pengolahan Awal Data	54
3.4.3	Model/Metode yang Diusulkan	54
3.4.4	Eksperimen dan Pengujian Model	56
3.4.5	Evaluasi dan Validasi Hasil	59
3.5	Jadwal Penelitian	60

BAB IV PEMBAHASAN HASIL PENELITIAN.....	63
4.1 Hasil Penelitian	63
4.1.1 Klasifikasi dengan Algoritma C4.5	63
4.1.2 Perhitungan pemilihan variable (Algoritma Genetika)	66
4.1.3 Pengujian Model dengan <i>10 Fold Cross Validation</i>	72
4.1.4 Eksperimen terhadap Indikator Model	74
4.1.5 Optimasi Model C4.5 dengan Algoritma Genetika	76
4.1.6 Optimasi Model C4.5 dengan Algoritma <i>PSO</i>	79
4.2 Evaluasi dan Validasi	82
4.2.1 <i>Confusion Matrix</i> dan <i>ROC Curve</i> Algoritma C4.5 dioptimasi dengan Algoritma Genetika	82
4.2.2 <i>Confusion Matrix</i> dan <i>ROC Curve</i> Algoritma C4.5 dioptimasi dengan Algoritma <i>PSO</i>	83
4.3 Hasil Komparasi	84
4.4 Penerapan Algoritma Terpilih	88
4.5 Implikasi Penelitian	90
4.5.1 Aspek Sistem	90
4.5.2 Aspek Manajerial	90
4.5.3 Aspek Penelitian Lanjut	90
BAB V PENUTUP.....	92
5.1 Kesimpulan.....	92
5.2 Saran.....	92
DAFTAR PUSTAKA	94
LAMPIRAN-LAMPIRAN.....	98
DAFTAR RIWAYAT HIDUP	136

DAFTAR GAMBAR

Gambar	Halaman
II- 1 Pohon keputusan dengan cabang dari atribut A5 tidak konsisten.....	16
II- 2 Pohon keputusan setelah di pruning.....	17
II- 3 Model yang diusulkan Yaoying Huang, et.al.....	38
II- 4 Model yang diusulkan Afzan Adam dan Khairuddin Omar	41
II- 5 Model yang diusulkan M. Lundin, et.al	43
II- 6 Kerangka pemikiran pemecahan masalah	47
III- 1 Langkah-langkah penelitian.....	53
III- 2 Model yang diusulkan.....	56
III- 3 Model yang diusulkan pada rapid mider 5.....	57
III- 4 Cross validation pada rapid miner 5	57
III- 5 Metode pada cross validation pada rapid miner 5	58
III- 6 Parameter untuk proses <i>decision tree</i>	59
IV- 1 Percabangan pertama.....	65
IV- 2 Model C4.5 tanpa dioptimasi.....	75
IV- 3 Model yang sudah dioptimasi dengan Algoritma Genetika	78
IV- 4 Model yang sudah dioptimasi dengan algoritma <i>PSO</i>	81
IV- 5 <i>ROC Curve</i> optimasi C4.5+AG	82
IV- 6 <i>ROC Curve</i> optimasi C4.5+ <i>PSO</i>	83
IV- 7 <i>Graphical User Interface (GUI)</i> Aplikasi Software Prediksi Kanker Payudara.....	88

DAFTAR TABEL

Tabel	Halaman
II- 1 Data penyakit jantung	17
II- 2 Model confusion matrix	31
II- 3 Komparasi antara model peneliti dengan yang lain	39
II- 4 Hasil klasifikasi bagian fitur berbeda.....	40
II- 5 Hasil Komparasi.....	43
II- 6 Rangkuman tinjauan studi.....	44
III- 1 Data awal yang belum diolah.....	51
III- 2 Spesifikasi komputer yang digunakan	59
III- 3 Jadwal penelitian.....	60
IV- 1 Sampel Dataset yang akan digunakan	63
IV- 2 Hasil nilai Info dan Gain dari atribut.....	64
IV- 3 Tabel penambahan atribut Class.....	66
IV- 4 Aplikasi Model terhadap data yang diuji.....	73
IV- 5 <i>Confusion Matrix</i> dari <i>pengujian model</i>	73
IV- 6 Indikator pohon keputusan	74
IV- 7 Indikator untuk Algoritma Genetika.....	76
IV- 8 Bobot Atribut terbaik dari Algoritma Genetika.....	77
IV- 9 Indikator untuk Algoritma PSO.....	79
IV- 10 Bobot Atribut terbaik dari Algoritma PSO	80
IV- 11 <i>Confusion Matrix</i> dari <i>Algoritma C4.5 + AG</i>	82
IV- 12 <i>Confusion Matrix</i> dari <i>Algoritma C4.5 + PSO</i>	83
IV- 13 Perbandingan model C4.5 sebelum dan sesudah optimasi	84
IV- 14 Metric of Software Quality Assurance (SQA)	88
IV- 15 Hasil Evaluasi <i>Software Quality Assurance (SQA)</i>	89
IV- 16 Spesifikasi <i>Hardware</i> dan <i>Software</i>	90

DAFTAR LAMPIRAN

Lampiran	Halaman
1 Dataset kanker payudara	98
2 Angket Penelitian	134



DAFTAR PUSTAKA

- [Adam 2006] Adam, Afzan, and Khairuddin Omar. 2006. "Computerized Breast Cancer Diagnosis with Genetic Algorithms and Neural Network." *Department of System and Management Sciences, Faculty of Technology and Information Sciences (Universiti Kebangsaan Malaysia)* 183-188.
- [Ali 2009] Ali, Amna, and Ali Tufail. 2009. "A Survey of Prediction Models for Breast Cancer Survivability." *Proceeding ICIS '09 Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human* 1259-1262 .
- [Alpaydin 2010] Alpaydin, Ethem. 2010. *Introduction to Machine Learning 2nd*. London: The MIT Press.
- [Anwar, 2014] Anwar H., "Variabel Penelitian", www.statistikian.blogspot.com (Diakses 12 Juni 2014).
- [Bellachia 2000] Bellachia, A. 2000. "Predicting Breast Cancer Survivability Using Data Mining Techniques." *Ninth Workshop on Mining Scientific and Engineering Datasets in conjunction with the Sixth SIAM International Conference on Data Mining (SDM 2006), Saturday, April 22, 2006*.
- [Bramer 2007] Bramer, Max. 2007. *Principles of Data Mining*. London: Springer.
- [Calle 2004] Calle, J. 2004. *Breast cancer facts and figures 2003—2004*. American Cancer Society. p. 1—27.
- [Dasarpendidikan 2014] Dasarpendidikan, "Pengertian Data Mining dan Metode Data", www.dasar-pendidikan.blogspot.com (Diakses 12 Juni 2014).
- [Dellen 2004] Dellen, Dursun, Glen WaLker, and Amit Kadam. 2004. "Predicting breast cancer survivability: a comparison of three data mining methods."
- [Gorunescu 2011] Gorunescu, Florin. 2011. *Data Mining Concepts, Models and Techniques*. Springer; 1st Edition. edition (June 24, 2011).
- [Hall 2011] Hall, T., S. Beecham, D. Bowes, Gray D., and S. Counsell. 2011. "A Systematic Literature Review on Fault Prediction Performance in Software Engineering."