

**OPTIMALISASI KINERJA HASIL KOMPARASI
ALGORITMA C4.5 DAN NAIVE BAYES MENGGUNAKAN
PARTICLE SWARM OPTIMIZATION :
STUDI KASUS PREDIKSI KELULUSAN UJIAN NASIONAL SMAN 80
JAKARTA**

TESIS



Oleh:

NASRULLOH ISNAIN

1211600851

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

JAKARTA

2014



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

LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : Nasrulloh Isnain

NIM : 1211600851

Program Studi : Magister Ilmu Komputer

Program : Strata Dua (S2)

Menyatakan bahwa TESIS yang berjudul :

**“OPTIMALISASI KINERJA HASIL KOMPARASI ALGORITMA C4.5
DAN NAIVE BAYES MENGGUNAKAN PARTICLE SWARM
OPTIMIZATION : STUDI KASUS PREDIKSI KELULUSAN UJIAN
NASIONAL SMAN 80 JAKARTA”**

1. Merupakan hasil karya tulis ilmiah sendiri dan bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademi 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, 25 Agustus 2014


(Nasrulloh Isnain)



PROGRAM STUDI MAGISTER ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PENGESAHAN

Nama Mahasiswa : Nasrulloh Isnain
NIM : 1211600851
Konsentrasi : Teknologi Sistem Informasi
Jenjang Studi : Strata 2
Judul : Optimalisasi Kinerja Hasil Komparasi Algoritma
C4.5 dan Naïve Bayes Menggunakan Particle Swarm
Optimization : Studi Kasus Prediksi Kelulusan Ujian
Nasional SMAN 80 Jakarta

Jakarta, 24 Agustus 2014

Tim Penguji

Ketua,

Dr. Moedjiono, M.Sc

Anggota,

Dr. Ir. Nazori AZ, M.T

Anggota,

Ir. Dana Indra Sensuse, MLIS, Ph.D.

Pembimbing,

Ir. Dana Indra Sensuse, MLIS, Ph.D.

Tanda Tangan

Ketua Program Studi

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

ABSTRAK

Didalam *data mining* terdapat teknik komparasi, dimana diujikan perbandingan antara algoritma satu dengan algoritma yang lain. Salah satu algoritma yang dapat digunakan dalam teknik komparasi adalah algoritma *C4.5* dan *Naïve Bayes*. Dari penelitian sebelumnya hasil dari teknik komparasi *C4.5* maupun *Naïve Bayes* didalam memprediksi suatu keakurasian dirasa belum optimal, sehingga diperlukan algoritma baru untuk lebih meningkatkan keakurasian dari algoritma sebelumnya dengan menggunakan pendekatan teknik optimasi, salah satunya adalah algoritma *Particle Swarm Optimization* (PSO). Dalam penelitian ini algoritma yang digunakan dalam teknik komparasi adalah algoritma *C4.5* dengan *Naïve Bayes*, dimana hasil dari teknik komparasi yang akurasinya lebih tinggi akan dioptimalkan dengan menggunakan algoritma PSO untuk prediksi kelulusan Ujian Nasional siswa. Hasil pengujian dengan mengukur kinerja kedua algoritma tersebut menggunakan metode pengujian *Confusion Matrix*, *Precision* dan *Recall*, diketahui bahwa algoritma *Naïve Bayes* memiliki nilai *accuracy* tertinggi, yaitu 93.46%, diikuti oleh metode *C4.5* dengan *accuracy* sebesar 91.33%. Dari algoritma yang paling tinggi yaitu algoritma *Naïve Bayes* dioptimalkan menggunakan algoritma PSO hasilnya menaikkan nilai *accuracy* menjadi 94.14% .

Kata kunci : *Data Mining*, *C4.5*, *Naïve Bayes*, *Particle Swarm Optimization*, komparasi, optimasi.

ABSTRACT

In data mining, there is comparison technique which its comparison between one algorithm and other one will be tested. Some of the algorithm which can be used in comparison technique are C4.5 and Naïve Bayes. From previous research, we know that the result of C4.5 or Naïve Bayes comparison technique doesn't show significant answer in predicting the accuracy. So, we need to find new algorithm that has better accuracy by using optimization approach, and one of them is Particle Swarm Optimization (PSO). The algorithm that will be used in comparison technique on this research are C4.5 and Naïve Bayes, which result of this comparison technique that has better accuracy will be optimized by using PSO algorithm to predict national exam passing grade for science-senior high school students after its accuracy is optimized by using Particle Swarm Optimization algorithm. The result of the best that we get after measuring both of the algorithm using confusion matrix testing method, precision and recall, whow that Naïve Bayes algorithm has the best accuracy, 93.46%, followed by C4.5 method that has 91.33% accuracy score. Algorithm with the best accuracy, Naïve Bayes, is optimized using PSO Algorithm, and the result shows that the accuracy score is increased to 94.14%.

Keywords: *Data Mining, C4.5, Naïve Bayes, Particle Swarm Optimization, comparison, optimized.*

KATA PENGANTAR

Puji dan syukur kehadiran Allah SWT dan junjungan besar Nabi Muhammad SAW atas berkah, rahmat dan karunia-Nya, sehingga penulisan tesis yang berjudul “*OPTIMALISASI KINERJA HASIL KOMPARASI ALGORITMA C4.5 DAN NAIVE BAYES MENGGUNAKAN PARTICLE SWARM OPTIMIZATION : STUDI KASUS PREDIKSI KELULUSAN UJIAN NASIONAL SMAN 80 JAKARTA*” dapat terselesaikan tepat pada waktunya walaupun banyak hambatan dalam penulisan tesis ini.

Tujuan pembuatan tesis ini adalah untuk memenuhi salah satu syarat dalam menyelesaikan jenjang Strata Dua (S-2) pada Program Studi Magister Ilmu Komputer (MKOM) Program Pascasarjana Universitas Budi Luhur. Pada kesempatan ini, penulis ingin menyampaikan ucapan terimakasih kepada :

1. Bapak Dr. Moedjiono, M.Sc selaku Direktur Program Pascasarjana Universitas Budi Luhur.
2. Bapak Dr. Ir. Nazori AZ, M.T selaku Ketua Program Studi Magister Komputer Universitas Budi Luhur.
3. Bapak Ir. Dana Indra Sensuse, MLIS, Ph.D. sebagai pembimbing tesis ini sehingga tesis ini dapat selesai pada waktunya.
4. Bapak Prof. Dr. H. Sumaryoto selaku Rektor Universitas Indraprasta PGRI, yang telah memberikan kesempatan untuk bisa melanjutkan pendidikan keningkat Strata Dua.
5. Bapak Kusnyoto, S.Pd selaku kepala sekolah, beserta seluruh staf Tata Usaha SMAN 80 Jakarta yang telah mengizinkan penulis melakukan riset untuk mendapatkan data atau informasi yang penulis butuhkan.
6. Orang tua tercinta Nur sodik dan Siti Maswiyah serta seluruh keluarga yang telah memberikan dukungan material dan moral kepada penulis.
7. Istri tercinta Rahmatika dan anak Amiera Khalilah Isnain, yang selalu memberikan motivasi dan semangat serta doa setiap saat kepada penulis untuk menyelesaikan studi.

8. Seluruh staf pengajar (dosen) Program Pascasarjana Magister Ilmu Komputer Universitas Budi Luhur yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
9. Seluruh staf dan karyawan Program Pascasarjana Magister Ilmu Komputer Universitas Budi Luhur yang telah melayani penulis dengan baik selama kuliah.
10. Seluruh teman-teman kuliah Program Pascasarjana Magister Ilmu Komputer Universitas Budi Luhur dan Universitas Indraprasta PGRI khususnya Heru Sulistiono, Muhamad Irsan dan Saputra Dwi Nurcahya.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini. Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 25 Agustus 2014

Nasrulloh Isnain

Penulis

DAFTAR ISI

	Halaman
HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
LEMBAR PERNYATAAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	2
1.2.1 Identifikasi Masalah	2
1.2.2 Batasan Masalah	2
1.2.3 Rumusan Masalah	2
1.3 Tujuan Dan Manfaat Penelitian	3
1.3.1 Tujuan Penelitian	3
1.3.2 Manfaat Penelitian	3
1.4 Sistematika Penulisan	3
BAB II LANDASAN TEORI DAN KERANGKA PEMIKIRAN	5
2.1 Tinjauan Pustaka	5
2.1.1 <i>Data Mining</i>	5
2.1.2 Teknik-teknik <i>data mining</i>	8
2.1.2.1 <i>Association Rule Mining</i>	8
2.1.2.2 <i>Classification</i>	9
2.1.2.3 <i>Clustering</i>	10

2.1.3	Algoritma C4.5	11
2.1.3.1	Konsep Algoritma C4.5	11
2.1.4	Algoritma <i>Naives Bayes</i>	12
2.1.4.1	Konsep Algoritma <i>Naives Bayes</i>	12
2.1.5	Algoritma <i>Particle Swarm Optimization</i> (PSO)	14
2.1.5.1	Konsep Algoritma <i>Particle Swarm Optimization</i> (PSO)	14
2.1.5.2	Proses Algoritma <i>Particle Swarm Optimization</i> (PSO)	15
2.1.6	Evaluasi dan Validasi Metode Klasifikasi <i>Data Mining</i>	16
2.1.7	Rapidminer	18
2.2	Tinjauan Studi yang Relevan	19
2.3	Tinjauan Objek Penelitian	26
2.3.1	Sejarah singkat berdirinya SMAN 80 Jakarta	26
2.3.2	Visi dan Misi SMAN 80 Jakarta	26
2.3.2.1	Visi SMAN 80 Jakarta	26
2.3.2.2	Misi SMAN 80 Jakarta	26
2.3.3	Struktur organisasi SMAN 80 Jakarta	27
2.3.4	Sarana dan Prasarana SI Pendukung Penelitian	30
2.3.4.1	Perangkat Keras (<i>Hardware</i>)	30
2.3.4.2	Perangkat Lunak (<i>Software</i>)	30
2.3.5	Dukungan SI Saat Ini	31
2.4	Kerangka Pemikiran	31
2.5	Hipotesis	33
BAB III METODOLOGI PENELITIAN		34
3.1	Jenis Penelitian	34
3.2	Populasi dan Sampel	34
3.2.1	Populasi	34
3.2.2	Sampel	34
3.3	Variabel Penelitian	35
3.4	Metode Pengumpulan Data	37
3.4.1	Studi Kepustakaan	37
3.4.2	Data Sekunder	37
3.5	Teknik Analisis Data	37

3.5.1	<i>Developing an understanding of the application domain</i>	37
3.5.2	<i>Selecting and creating</i>	38
3.5.3	<i>Preprocessing and cleansing</i>	38
3.5.4	<i>Data Transformation</i>	38
3.5.5	<i>Choosing the appropriate data mining task</i>	40
3.5.6	<i>Choosing the data mining algorithm</i>	40
3.5.7	<i>Employing the data mining algorithm</i>	41
3.5.8	<i>Evaluation</i>	41
3.5.9	<i>Using the discovered knowledge</i>	41
3.6	Langkah-langkah penelitian	42
3.7	Jadwal Penelitian	44
BAB IV PEMBAHASAN HASIL PENELITIAN		46
4.1	Pengukuran Penelitian	46
4.1.1	Pengelompokan dan Analisis Data	46
4.1.2	Algoritma C4.5	46
4.1.3	Algoritma Naïve Bayes	50
4.2	Evaluasi dan Validasi	52
4.2.1	<i>Confusion Matrix</i> Algoritma C4.5	52
4.2.2	<i>Confusion Matrix</i> Algoritma Naïve Bayes	53
4.3	Analisa Hasil Komparasi	54
4.4	Optimasi Algoritma PSO dari Algoritma Terpilih	54
4.5	Penerapan Algoritma PSO	56
4.6	Implikasi Penelitian	58
4.6.1	Aspek Manajerial	58
4.6.2	Aspek Sistem	58
4.6.3	Rencana Implementasi	58
BAB V PENUTUP		60
5.1	Kesimpulan	60
5.2	Saran	60

DAFTAR PUSTAKA	61
LAMPIRAN-LAMPIRAN	63
DAFTAR RIWAYAT HIDUP	81



DAFTAR GAMBAR

2.1	Tahap-tahap <i>data mining</i>	6
2.2	Langkah-langkah dalam Proses <i>Data Mining</i>	8
2.3	PosteriorProbability	13
2.4	Tampilan layar <i>Rapid Miner</i>	19
2.5	Struktur Organisasi SMAN 80 Jakarta	27
2.6	Website SMAN 80 Jakarta.....	31
2.7	Kerangka Pemikiran.....	32
3.1	Langkah-langkah penelitian.....	42
4.1	Decision Tree klasifikasi UN siswa dengan C4.5.....	49
4.2	<i>Text View Model confusion matrix</i> untuk algoritma C4.5	52
4.3	<i>Text View Model confusion matrix</i> untuk algoritma Naïve Bayes	53
4.4	<i>Text View Model confusion matrix</i> algoritma Naïve Bayes+PSO	55
4.5	Aplikasi untuk klasifikasi prediksi kelulusan UN	57

DAFTAR TABEL

2.1	Model Confusion Matrix.....	17
2.2	Tinjauan studi terdahulu yang relevan	24
3.1	Atribut nilai dan kategori	36
3.2	Data siswa	38
3.3	Data siswa <i>processing, cleansing</i> dan transformasi data	39
3.4	Jadwal Penelitian	44
4.1	Hasil perhitungan entrophy dan gain menentukan simpul akar	47
4.2	Perhitungan probabilitas prior	50
4.3	Perhitungan menentukan klasifikasi data testing x	51
4.4	Model tabel confusion matrix algoritma C4.5	52
4.5	Model tabel confusion matrix algoritma Naïve Bayes	53
4.6	Komparasi accuracy, precision, recall C4.5 dan Naïve Bayes	54
4.7	Model tabel confusion matrix algoritma Naïve Bayes + PSO.....	55
4.8	Komparasi accuracy, precision, recall Naïve Bayes + PSO	56
4.9	Data baru untuk penerapan algoritma terpilih	56
4.10	Confusion Matrix data baru algoritma Naïve Bayes.....	57
4.11	Rencana implementasi sistem.....	58

DAFTAR PUSTAKA

- [Ariani dkk] Ariani, D., Fahriza, A., Prasetyaningrum, I. "Optimasi Penjadwalan Mata Kuliah Di Jurusan Teknik Informatika Pens Dengan Menggunakan Algoritma Particle Swarm Optimization (PSO)". Jurnal, PENS-ITS Keputih Sukolilo Surabaya.
- [Bramer 2007] Bramer, Max. 2007. *Principles of Data Mining*. London: Springer.
- [Deker dkk 2009] Decker, G. W., Pechenizkiy, M., Vleeshouwers, J. M., 2009. "Predicting Students Drop Out: A Case Study". *Educational Data Mining*.
- [Dhika 2012] Dhika, Harry. "Kajian Komparasi Penerapan Algoritma C4.5, Naïve Bayes, Dan Neural Network Dalam Pemilihan Mitra Kerja Penyedia Jasa Transportasi: Studi Kasus Cv. Viradi Global Pratama". Tesis, STIMIK Nusa Mandiri.
- [Gorunescu 2011] Gorunescu, F. 2011. *Data Mining Concepts ,Models and Techniques* (Intelligent Systems Reference Library Volume 12). Berlin: Springer.
- [Haupt & Haupt 2004] Haupt, R. L., Haupt, S.E. *Practical Genetic Algorithm* (second edition). New Jersey : Wiley.
- [Han 2007] Han, Jiawei. Pei, Jian. 2007. "Mining Frequent Pattern by Pattern Growth: Metodology and Implication". Diakses pada Juni 2014 dari <http://www.acm.org/sigs/sigkdd/explorations/issues/2-2-2000-12/han.pdf>.
- [Han & Kamber 2006] Han, J. & Kamber, M. 2006. *Data Mining Concept and Techniques*. San Fransisco: Morgan Kauffman.
- [Irfansyah 2013] Irfansyah, Puput. "Kajian Komparasi Penerapan Algorithma Data Mining (C4.5, Bayessian Classifier, Dan Neural Network) Dalam Menentukan Promosi Jabatan: Studi Kasus PT. Intisel Prodaktifakom". Tesis, Universitas Budi Luhur.
- [Kusrini 2009] Kusrini, E.T. Luthfi. *Algoritma Data Mining*. Yogyakarta : Andi Offset.
- [Larose 2005] Larose, Daniel. T. *Discovering Knowledge in Data: An Introduction to Data mining*. New Jersey : Jhon Willey & Son Inc.
- [Liao 2007] Liao, T. W. & Triantaphyllou, E. 2007. "Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications". (Series on