

**IMPLEMENTASI ALGORITMA DECISION TREE ID3 UNTUK
KLASIFIKASI ARTIKEL BERITA ONLINE**

Tugas Akhir



Oleh :

Andrew Junior Ongi Karyanto

22 07 4218

**Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Tahun 2011**

**IMPLEMENTASI ALGORITMA DECISION TREE ID3 UNTUK
KLASIFIKASI ARTIKEL BERITA ONLINE**

Tugas Akhir



**Diajukan kepada Fakultas Teknik Informatika
Universitas Kristen Duta Wacana
Sebagai salah satu syarat dalam memperoleh gelar
Sarjana Komputer**



**Disusun oleh :
Andrew Junior Ongi Karyanto
22 07 4218**

**Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Tahun 2011**

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul :
Implementasi Algoritma Decision Tree ID3 untuk Klasifikasi Artikel Berita
Online

Yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana
Komputer pada pendidikan sarjana Program Studi Teknik Informatika, Fakultas
Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan
atau duplikasi dari skripsi kesarjanaaan di lingkungan Universitas Kristen Duta
Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang
sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau
tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar
kesarjanaaan saya.

Yogyakarta, 20 Mei 2011



(Andrew Junior Ongi Karyanto)

22 07 4218

HALAMAN PERSETUJUAN

Judul : Implementasi Algoritma Decision Tree ID3 untuk
Klasifikasi Artikel Berita Online
Nama : Andrew Junior Ongi Karyanto
NIM : 22 07 4218
Mata Kuliah : Tugas Akhir
Kode : TI2126
Semester : Genap
Tahun akademik : 2010/2011

Telah diperiksa dan disetujui
Di Yogyakarta,
Pada tanggal 20 Mei 2011



Dosen Pembimbing I

A handwritten signature in blue ink, appearing to read 'Lucia', is written over the name.

Lucia Dwi Krisnawati, S.S., M.A.

Dosen Pembimbing II

A handwritten signature in blue ink, appearing to read 'Antonius', is written over the name.

Antonius Rachmat C, S.Kom., M.Cs.

HALAMAN PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA DECISION TREE ID3 UNTUK
KLASIFIKASI ARTIKEL BERITA ONLINE

Oleh : Andrew Junior Ongi Karyanto / 22 07 4218

Dipertahankan di depan dewan Penguji Tugas Akhir
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu

Syarat memperoleh gelar

Sarjana Komputer

Pada tanggal

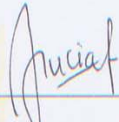



14 Juni 2011

Yogyakarta, 20/6/2011

Mengesahkan,

Dewan Penguji:

1. Lucia D. Krishawati, S.S., M.A.
2. Antonius Rachmat, S.Kom., M.Cs.
3. Ir. Sri Suwarno, M.Eng
4. Yuan Lukito, S.Kom.

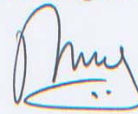





Dekan



(Drs. Wimmie Handiwidjojo, MIT)

Ketua Program Studi



(Nugroho Agus Haryono S.Si.,MSi.)

UCAPAN TERIMA KASIH

Puji syukur kepada Tuhan atas segala berkat, bimbingan, dan perlindungan-Nya sehingga penulis menyelesaikan Tugas Akhir dengan judul Implementasi Algoritma Decision Tree ID3 untuk Klasifikasi Artikel Berita Online.

Penulisan laporan ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu bertujuan melatih mahasiswa untuk dapat menghasilkan suatu karya yang dapat dipertanggungjawabkan secara ilmiah, sehingga dapat bermanfaat bagi penggunanya.

Dalam menyelesaikan pembuatan program dan laporan Tugas Akhir ini, penulis telah banyak menerima bimbingan, saran dan masukan dari berbagai pihak, baik secara langsung maupun secara tidak langsung. Untuk itu dengan segala kerendahan hati, pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada :

1. **Lucia D. Krisnawati, S.S, M.A.** selaku dosen pembimbing I yang telah memberikan bimbingannya dengan sabar dan baik kepada penulis, juga kepada
2. **Antonius Rachmat, S.Kom, M.Cs.** selaku dosen pembimbing II atas bimbingan, petunjuk dan masukan yang diberikan selama pengerjaan tugas ini sejak awal hingga akhir.
3. Keluarga tercinta yang selalu memberi dukungan dan do'a bagi penulis.
4. Rekan-rekan dan pihak-pihak yang tidak dapat penulis sebutkan satu persatu yang secara langsung maupun tidak langsung yang telah mendukung penyelesaian tugas ini. Terima kasih atas dukungan dan do'anya.

5. Yang terakhir dan yang terpenting, kepada Tuhan Yesus Kristus, yang memberikan seluruh kekuatan, waktu, dan kesehatan selama penulis mengerjakan Tugas Akhir ini.

Penulis menyadari bahwa laporan Tugas Akhir yang penulis susun ini masih banyak kekurangannya. Oleh karena itu, penulis mohon saran dan kritik dari pembaca guna kesempurnaan tulisan ini. Sehingga suatu saat penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata penulis ingin meminta maaf bila ada kesalahan baik dalam penyusunan laporan maupun yang pernah penulis lakukan sewaktu membuat program Tugas Akhir. Sekali-lagi penulis mohon maaf yang sebesar-besarnya. Dan semoga ini dapat berguna bagi kita semua.

Yogyakarta, Mei 2011

Penulis



© UKDWN

INTISARI

Implementasi Algoritma Decision Tree ID3 untuk Klasifikasi Artikel Berita Online

Perkembangan teknologi telah membuat banyak informasi bermunculan. Pada saat ini, media informasi tersebut tidak hanya dalam bentuk kertas saja namun dokumen digital. Untuk mengorganisir dokumen digital tersebut juga cukup membutuhkan waktu. Salah satu cara agar bisa mengorganisir informasi-informasi tersebut dengan mudah adalah dengan melakukan klasifikasi dokumen secara otomatis, sehingga dokumen-dokumen tersebut dapat memiliki kategori secara otomatis sesuai kategori yang sudah dipersiapkan.

Metode klasifikasi yang akan diterapkan adalah metode *Decision Tree ID3*. Metode *Decision Tree ID3* menggunakan pohon keputusan untuk menghasilkan aturan-aturan yang digunakan untuk klasifikasi data. Dalam penelitian ini, kelas yang disiapkan untuk implementasi metode *Decision Tree ID3* ini adalah sepakbola, otomotif, raket, dan tinju. Sistem akan melakukan klasifikasi terhadap masukan yang berasal dari *RSS feed* berdasarkan pohon keputusan yang dihasilkan dari metode *Decision Tree ID3*. Subyek utama dari penelitian ini adalah pada nilai akurasi yang dihasilkan oleh metode tersebut.

Beberapa kesimpulan yang dapat diperoleh setelah melakukan penelitian tentang implementasi algoritma *Decision Tree ID3* untuk klasifikasi artikel berita *online* adalah algoritma *Decision Tree ID3* dapat diterapkan untuk klasifikasi artikel berita *online* dan dalam uji coba sistem nilai akurasi tertinggi didapat dengan menggunakan *feature selection* bernilai 10% dan 15%. Meskipun tidak bisa 100% akurat, namun algoritma ini dapat menghasilkan nilai akurasi yang cukup tinggi, yaitu 97,05%.

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Berkembangnya teknologi telah membuat banyak informasi bermunculan. Informasi-informasi tersebut tertuang dalam bentuk dokumen terutama dokumen digital. Semakin banyak informasi yang ada maka semakin banyak dokumen-dokumen yang digunakan. Untuk bisa mengorganisir informasi-informasi tersebut dengan mudah, maka dibutuhkan klasifikasi dokumen secara otomatis.

Salah satu metode yang dapat digunakan untuk klasifikasi adalah metode klasifikasi dengan *Decision Tree*, di mana metode dari *Decision Tree* yang umum dipakai adalah ID3. *Decision Tree* adalah sebuah pohon di mana setiap node cabang merupakan pilihan antara sejumlah alternatif, dan setiap simpul daun merupakan keputusan. *Decision Tree* biasanya digunakan untuk memperoleh informasi untuk tujuan pengambilan keputusan. *Decision Tree* dimulai dengan *root node* yang digunakan untuk mengambil keputusan. Dari *node* ini, pengguna memisahkan tiap *node* secara rekursif menurut algoritma pelatihan *Decision Tree*. Hasil akhir adalah *Decision Tree* di mana tiap cabang merupakan skenario kemungkinan keputusan dan hasilnya

Dalam penelitian yang akan dilakukan sebagai tugas akhir ini, penulis akan membuat sistem klasifikasi artikel secara otomatis dengan metode *Decision Tree ID3*. Melalui program ini, artikel-artikel berita yang ada akan secara otomatis terklasifikasi sehingga mempermudah pengguna dalam melakukan pencarian artikel-artikel tersebut sesuai dengan kelas yang sudah ditentukan. Dari program yang akan dibuat ini, peneliti akan melakukan penelitian mengenai tingkat kinerja metode *Decision Tree* untuk klasifikasi artikel berita, sehingga dapat diketahui tingkat keakuratan metode yang digunakan untuk klasifikasi.

1.2 Rumusan Masalah

Dalam tugas akhir ini, penulis akan merancang dan membangun sistem yang akan melakukan proses klasifikasi terhadap suatu artikel online secara otomatis.

Rumusan masalah yang akan dibuat oleh sistem secara garis besar adalah sebagai berikut:

1. Bagaimana mengimplementasikan *Decision Tree ID3* untuk klasifikasi artikel online?
2. Berapa persen tingkat akurasi dari implementasi *Decision Tree ID3* untuk klasifikasi artikel online?
3. Bagaimana pengaruh *feature selection* terhadap tingkat akurasi implementasi *Decision Tree ID3* untuk klasifikasi artikel online?

1.3 Batasan Masalah

Untuk mencegah meluasnya masalah dalam tugas akhir ini, maka diberikan batasan sebagai berikut:

1. Artikel berita terbatas pada artikel olahraga, dan kelas-kelas untuk artikel tersebut dibagi menjadi: olahraga sepak bola, olahraga raket, olahraga otomotif, dan olahraga tinju.
2. Artikel yang digunakan adalah artikel berita olahraga yang berbahasa Indonesia.
3. Data pelatihan diambil dari sumber berita olahraga online yang berbahasa Indonesia, yaitu <http://www.sports.okezone.com>, <http://www.bola.okezone.com>, <http://www.detiksport.com>, dan <http://www.bolanews.com>, dengan jumlah minimal 200 berita.
4. Input data pelatihan berupa *file plain-text* (.txt), yang isinya adalah dari berita olahraga online tersebut yang di copy ke *file text* tersebut.

5. Artikel olahraga untuk data implementasi program adalah artikel berita olahraga yang diambil dari *RSS feed* yang disediakan okezone, yaitu: <http://sports.okezone.com/rss> dan <http://bola.okezone.com/rss>.
6. Tidak melakukan *stemming* untuk *token-token* yang ada pada sistem yang akan dibangun.

1.4 Hipotesis

Dari penelitian yang akan dilakukan ini, penulis memberikan hipotesis bahwa metode *Decision Tree ID3* dapat melakukan klasifikasi yang relevan untuk setiap artikel berita dengan tingkat akurasi di atas 60%.

1.5 Tujuan Penelitian

Tujuan dari penelitian yang akan dibangun adalah sebagai berikut:

1. Menerapkan metode *Decision Tree ID3* untuk membangun sistem klasifikasi secara otomatis.
2. Mengukur tingkat akurasi klasifikasi yang dihasilkan dari metode *Decision Tree ID3*.

1.6 Metodologi Penelitian

Metodologi yang digunakan pada penelitian ini adalah sebagai berikut:

1. Metode Pengumpulan Data

Untuk mendapatkan data-data yang diperlukan, pertama-tama penulis melakukan studi pustaka untuk memperdalam pengetahuan dalam

pembangunan sistem dengan cara mempelajari definisi, teori-teori, dan perhitungan dari buku atau jurnal yang berhubungan dengan metode *Decision Tree ID3*. Selain itu, penulis juga mencari data berupa dokumen *plain text* yang bisa didapatkan dari Internet dengan cara mengcopy dari situs-situs web, kemudian di *paste* ke dalam sebuah *file .txt* sebagai bahan pelatihan untuk pengembangan sistem.

2. Metode Pengembangan Sistem

Metode yang dipakai untuk mengembangkan sistem dalam penelitian ini adalah metode *Decision Tree ID3*. Metode ini digunakan sebagai instrumen dalam pembangunan program untuk melakukan klasifikasi secara otomatis.

3. Metode Evaluasi Sistem

Untuk mengevaluasi tingkat akurasi sistem yang dibangun, metode yang digunakan adalah dengan menghitung persentase ketepatan kelas artikel yang dibangun dengan sistem dengan kelas artikel yang disediakan oleh situs penyedia berita.

1.7 Sistematika Penulisan

Sistematika penulisan laporan kerja praktik ini dibagi menjadi 5 bagian utama. Bab pertama yaitu pendahuluan berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penulisan, batasan masalah, hipotesis, tujuan penelitian, metodologi penelitian, dan sistematika penulisan.

Bab kedua yaitu landasan teori, berisi tinjauan pustaka dan landasan teori-teori penunjang yang diperlukan untuk penulisan tugas akhir, yaitu berupa teori-teori *machine learning*, *data mining*, *text mining*, dan teori-teori yang khususnya berhubungan dengan *Decision Tree ID3*.

Bab ketiga yaitu analisis dan perancangan sistem, berisi kebutuhan *hardware* dan *software* sistem, spesifikasi sistem, arsitektur sistem, use case sistem dan perancangan dari sistem yang meliputi perancangan masukan, keluaran, proses, dan pengujian, dan perancangan basis data.

Bab keempat yaitu implementasi dan analisis sistem, berisi penjelasan dari implementasi sistem yang dibuat, meliputi penjabaran hasil penelitian, serta kelebihan dan kekurangan dari sistem yang dibangun.

Bab kelima yaitu kesimpulan dan saran, berisi kesimpulan dari tugas akhir ini dan saran pengembangan program pada masa yang akan datang.

© UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Setelah dilakukan penelitian mengenai penerapan algoritma *Decision Tree ID3* untuk klasifikasi artikel berita online khususnya berita olahraga penambahan nilai *feature selection* belum tentu meningkatkan nilai akurasi sistem. Sedangkan penambahan data sendiri cenderung meningkatkan nilai akurasi. Hal ini cenderung bergantung pada kualitas data pelatihan dan data masukan dari *RSS* itu sendiri. Kualitas data pelatihan akan sangat berpengaruh pada pohon keputusan yang terbentuk, sementara data masukan sistem akan sangat berpengaruh pada kategori-kelas yang dihasilkan sistem.

Data yang paling sering salah adalah kategori sepak bola menjadi kategori tinju (*feature selection* 5%), sedangkan untuk *feature selection* lainnya kategori sepakbola sering tertukar dengan kategori otomotif. Kesalahan ini mungkin disebabkan oleh miripnya kata kunci diantara kedua kelas tersebut, tidak seperti kategori tinju yang memiliki kata kunci yang tidak terdapat di kelas lainnya yaitu petinju, sehingga terjadi kesalahan karena pohon keputusan yang terbentuk juga salah.

Nilai *feature selection* yang menghasilkan nilai akurasi paling tinggi adalah nilai *feature selection* sebesar 10% dan 15%. Kedua nilai *feature selection* tersebut memberikan nilai akurasi yang sama besarnya sampai dengan pengujian yang ke-8 dengan 178 data olahraga dengan nilai akurasi 97,02%.

Berita olahraga dengan kategori tinju merupakan kategori yang memiliki nilai akurasi paling tinggi diantara kategori olahraga lainnya, yaitu dengan nilai 100% sampai dengan pengujian ke-8.

Dengan demikian dapat disimpulkan bahwa algoritma *Decision Tree ID3* merupakan algoritma yang cukup ideal untuk melakukan klasifikasi karena dapat menghasilkan nilai akurasi yang tinggi, yaitu sampai 97,02%.

5.2 Saran

Untuk pengembangan sistem selanjutnya yang lebih baik, penulis mempunyai beberapa saran, diantaranya adalah sumber berita yang lebih dari satu. Kategori berita yang akan diteliti sebaiknya mempunyai jumlah *update* yang cukup berimbang, tidak berat sebelah.

Sistem yang dibangun sebaiknya dapat melakukan *web crawler* sehingga mampu melakukan *grabbing* secara dinamis terhadap berbagai sumber berita yang ada dan sistem tidak terpengaruh oleh *layout* dari halaman *web* tersebut sehingga ketika terjadi perubahan layout, sistem masih mampu melakukan *grabbing* terhadap konten *web* tersebut.

Untuk pencocokan pohon keputusan yang telah dibangun dan data masukan dari RSS feed, sebaiknya diperhitungkan juga pengaruh frekuensi kemunculan dan bobot kata dari data masukan tersebut untuk meningkatkan hasil akurasi sistem. Kata yang muncul pada data masukan namun mungkin tidak berhubungan dengan isi berita tersebut tidak dicocokkan dengan aturan dari pohon keputusan meski kata tersebut terdapat dalam aturan tersebut.



DAFTAR PUSTAKA

- Bittner, K., & Spence, I. (2003). *Use case modeling*. Boston: Addison-Wesley.
- Cios, K.J., Pedrycz, W., Swiniarski, R.W., & Kurgan, L.A. (2007). *Data Mining A Knowledge Discovery Approach*. New York: Springer Science+Business Media.
- Defiyanti, S. (2009). *Perbandingan Kinerja Algoritma Id3 Dan C4.5 Dalam Klasifikasi Spam-Mail*. Diakses 25 Maret 2011, dari [http://openstorage.gunadarma.ac.id/~mwiryana/KOMMIT/per-artikel/03-02-004-Perbandingan\[Sofi\].pdf](http://openstorage.gunadarma.ac.id/~mwiryana/KOMMIT/per-artikel/03-02-004-Perbandingan[Sofi].pdf).
- Han, J., & Kamber, K. (2001). *Data Mining: Concepts and Technique*. San Fransisco: Morgan Kaufmann Publisher.
- Kecskemeti, G. (2004). *Comparison of Classification Methods by Using the Reuters Database*. Diakses 25 Maret 2011, dari <http://www.lpds.sztaki.hu/~kecskemeti/publications/mcclassify.pdf>.
- Ramadan, R. (2006). *Penerapan Pohon Untuk Klasifikasi Dokumen Teks Berbahasa Inggris*. Diakses 25 Maret 2011, dari <http://www.informatika.org/~rinaldi/Matdis/2006-2007/Makalah/Makalah0607-121.pdf>.

