

# **Implementasi Metode *Naïve Bayes* untuk Pengklasifikasian Dokumen Teks**

Tugas Akhir



Diajukan kepada

Fakultas Teknologi Informasi Program Studi Teknik Informatika

Universitas Kristen Duta Wacana

Sebagai salah satu syarat memperoleh gelar

Sarjana Komputer



Disusun Oleh :

Anggun Nur Aruma Nugroho

22002351

Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

2011

## PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul :

### IMPLEMENTASI METODE NAÏVE BAYES UNTUK PENGKLASIFIKASIAN DOKUMEN TEKS

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 kesarjanaan dilingkungan Universitas Kristen Duta Wacana atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagai mana mestinya.

Jika dikemudian hari didapati bahwa skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia menerima sanksi berupa pencabutan gelar kesarjanaan saya.

Yogyakarta, 25 Desember 2010



Anggun Nur Aruma Nugroho

22002351

## HALAMAN PERSETUJUAN

Judul : **Implementasi Metode Naïve Bayes untuk Pengklasifikasian  
Dokumen Teks**

Nama : Anggun Nur Aruma Nugroho

NIM : 22002351

Mata Kuliah : Tugas akhir

Kode : TI2126

Semester : Gasal

Tahun Akademik : 2010 / 2011


Telah diperiksa dan disetujui

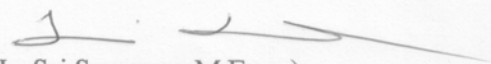
Di Yogyakarta

Pada Tanggal 17 Desember 2010

Dosen Pembimbing I

Dosen Pembimbing II

  
( Restyandito, S.Kom., M.SIS. )

  
( Ir. Sri Suwarno, M.Eng. )

HALAMAN PENGESAHAN

SKRIPSI

Implementasi Metode *Naïve Bayes* untuk Pengklasifikasian  
Dokumen Teks

Oleh : Anggun Nur Aruma Nugroho / 22002351

Dipertahankan di depan dewan penguji Tugas Akhir / Skripsi  
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

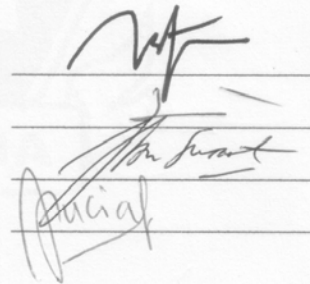
7-1- 2011

Yogyakarta, <sup>12</sup> Januari 2011

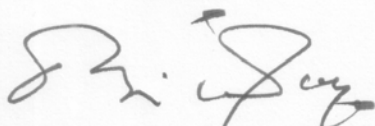
Mengesahkan,

Dewan Penguji :

1. Restyandito, S.Kom., M.SIS.
2. Ir. Sri Suwarno, M.Eng.
3. Budi Susanto, S.Kom., M.T.
4. Lucia Dwi Krisnawati, M.A.



Dekan



(Drs. Wimmie Handwidjojo, MIT.)

Ketua Program Studi



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

## UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepadaMu Ya Allah, atas segala nikmat yang telah Engkau berikan hingga saat ini. Sehingga penulis dapat merasakan anugrah proses penyelesaian Tugas Akhir selama ini.

Penulisan laporan Tugas Akhir ini merupakan kelengkapan dan pemenuhan dari salah satu syarat untuk 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 terima kasih kepada :

1. **Ibunda** tercinta, yang memaksa penulis untuk menyelesaikan Tugas Akhir sampai titik darah penghabisan.
2. **Mantan Pacar** tersayang, yang selalu mengerti dan mendukung.
3. **Bapak Budi Susanto** yang telah memberikan judul serta memberikan bimbingan dan petunjuk dengan sabar dan baik kepada penulis.
4. **Ibu Umi** atas kritik dan saran penulisan pada bimbingan sebelumnya.
5. **Bapak Restyandito** dan **Bapak Sri Suwarno**, yang telah mengizinkan penulis untuk maju pendadaran. Sekali lagi terima kasih.
6. **Mister Why** dan **Flasherman** atas semangat suka dan duka nya.
7. Teman teman **DWPh Club** yang luar biasa caci maki positif nya, terima kasih atas pengalaman pengalaman yang tak terbayarkan.
8. Pihak lain yang tidak dapat penulis sebutkan satu persatu, sehingga Tugas Akhir ini dapat terselasaikan dengan baik.

Penulis menyadari bahwa program dan laporan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari pembaca sekalian, sehingga suatu saat penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata penulis ingin meminta maaf bila ada kesalahan baik penyusunan laporan maupun yang penulis lakukan sewaktu membuat program Tugas Akhir ini. Semoga tugas akhir ini dapat berguna bagi kita semua.

Yogyakarta, 25 Desember 2010

Penulis



© UKDWN

## INTISARI

### Implementasi Metode *Naïve Bayes* untuk Pengklasifikasian Dokumen Teks

Penyaringan atau klasifikasi diperlukan untuk memilah-milah dokumen teks menjadi satu atau lebih kategori berdasarkan isinya. Sebagai contoh pada kasus *newsgroup* ataupun artikel, tentu akan sulit bagi kita untuk mengolah dan menentukan suatu dokumen termasuk dalam kategori yang mana jika terdapat ratusan bahkan ribuan dokumen. Serta akan memerlukan waktu yang lama bagi *user* untuk mengetahui intisari dari dokumen tersebut, jika *user* harus membaca halaman per halaman.

Pengklasifikasian dokumen teks adalah proses pemilihan dokumen ke dalam kategori tertentu berdasarkan isi dokumen. Berdasarkan latar belakang di atas, rumusan masalah yang akan dibahas dalam penelitian ini adalah :

1. Membangun sebuah program untuk mengklasifikasikan suatu dokumen teks berbahasa Inggris dengan menggunakan algoritma Naive Bayes.
2. Menerapkan Algoritma Porter Stemming untuk mengubah semua kata dalam dokumen teks menjadi bentuk dasarnya pada tahap *preprocessing*.

Hasil pengklasifikasian yang dilakukan bergantung pada inputan yang dimasukkan. Semakin besar/ banyak data/ kata yang diinputkan, maka akan semakin akurat hasil pengklasifikasian.

## DAFTAR ISI

HALAMAN JUDUL.....	i
PERNYATAAN KEASLIAN TUGAS AKHIR.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
UCAPAN TERIMA KASIH.....	v
INTISARI.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	2
1.5. Metode.....	3
1.6. Sistematika Penulisan.....	3
TINJAUAN PUSTAKA.....	5
2.1. Tinjauan Pustaka.....	5
2.2. Landasan Teori.....	10
PERANCANGAN SISTEM.....	12
3.1. Perancangan Input.....	12
3.2. Kamus Data.....	18
3.3. Antarmuka Pengguna.....	20
IMPLEMENTASI DAN ANALISA SISTEM.....	21
4.1. Perancangan Sistem.....	21
4.2. Pengujian Sistem.....	30



KESIMPULAN DAN SARAN.....	39
5.1. Kesimpulan .....	39
5.2. Saran.....	39
DAFTAR PUSTAKA .....	40

© UKDW

## DAFTAR TABEL

Tabel 3.1.....	14
Tabel 3.2.....	18
Tabel 3.3.....	18
Tabel 3.4.....	19
Tabel 4.1.....	32
Tabel 4.2.....	33
Tabel 4.3.....	34
Tabel 4.4.....	35
Tabel 4.5.....	36
Tabel 4.6.....	37



UKDWN

## DAFTAR GAMBAR

Gambar 2.1 .....	7
Gambar 3.1 .....	12
Gambar 3.2 .....	12
Gambar 3.3 .....	15
Gambar 3.4 .....	16
Gambar 3.5 .....	17
Gambar 3.6 .....	18
Gambar 4.1 .....	27
Gambar 4.2 .....	27
Gambar 4.3 .....	29
Gambar 4.4 .....	31
Gambar 4.5 .....	3



UKDIN

## **BAB 5**

### **PENUTUP**

#### **5.1. Kesimpulan**

Berdasarkan perancangan dan implementasi program, maka diperoleh kesimpulan sebagai berikut :

- Proses klasifikasi dapat diterapkan pada sebuah dokumen teks dengan metode *TF-IDF* dan Naive-Bayesian. *TF-IDF* digunakan untuk pembobotan kata dan Naive-Bayesian dipakai untuk proses klasifikasi dokumen tersebut berdasarkan rata-rata bobot kata yang dimiliki tiap dokumen.
- Jumlah persentase *threshold* yang diinputkan sangat mempengaruhi hasil perhitungan probabilitas, semakin besar maka akan semakin meleset *output* yang dihasilkan.
- Hasil klasifikasi menjadi tidak akurat apabila dokumen yang diinputkan tidak memiliki cukup banyak jenis kata.

#### **5.2. Saran**

Saran untuk mengembangkan sistem peringkasan dokumen teks ini lebih lanjut adalah sebagai berikut :

- Program ini masih belum berjalan sesuai teori, dapat dikembangkan dengan menambah dan atau mengganti rumus probabilitas pada tahap *processingnya* agar hasil lebih akurat.
- Waktu yang diperlukan sistem untuk melakukan proses pengklasifikasian dokumen teks sangat tergantung dari banyaknya data yang diinputkan.
- Perlu adanya pengembangan untuk mempercepat waktu kerja sistem, dapat menggunakan cara *desktop programming*.

## DAFTAR PUSTAKA

- Berry, Michael J. A. & Linoff, Gordon S., *Mastering Data Mining*, John Wiley & Sons, 2000
- Damerau Fred J, Sholom M Weiss, Nitin Indurkha, Tong Zhang. *Text Mining : Predictive Methods for Analyzing Unstructure Information*. Springer Science, United States, 2005
- Han, Jiawei & Kamber, Micheline, *Data Mining : Concept and Techniques*, Morgan Kauffman, 2001
- Kantardzic, Mehmed, *Data Mining Concepts, Models, Methods, and Algorithms*, John Wiley & Sons, 2003

## REFERENCE

[1] Jiawei Han dan Micheline Kamber, *Data Mining : Concept and Techniques*, San Francisco: Morgan Kauffman Publishers, 2001, hal: 296

[2] *Ibid.*

