

**IMPLEMENTASI METODE FUZZY K-NEAREST NEIGHBOR
DALAM KLASIFIKASI TUGAS AKHIR SEBAGAI PENENTUAN
BIDANG MINAT**

Studi Kasus Dokumen Abstraksi Tugas Akhir Fakultas Teknologi Informasi
Program Studi Teknik Informatika

Tugas Akhir



Oleh

Bernadetha Ayu Setyaningtyas

22074226

Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

Tahun 2011

**IMPLEMENTASI METODE FUZZY K-NEAREST NEIGHBOR
DALAM KLASIFIKASI TUGAS AKHIR SEBAGAI PENENTUAN
BIDANG MINAT**

Studi Kasus Dokumen Abstraksi Tugas Akhir Fakultas Teknologi Informasi
Program Studi Teknik Informatika

Tugas Akhir



©
Diajukan kepada Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai salah satu syarat dalam memperoleh gelar
Sarjana Teknik Informatika

Disusun oleh:

Bernadetha Ayu Setianingtyas

22074226

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 Metode Fuzzy K-Nearest Neighbor Dalam Klasifikasi Tugas Akhir Sebagai Penentuan Bidang Minat

Yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Teknik Informatika pada pendidikan Sarjana Program Studi Teknik Informatikan, Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian 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 kesarjanaan saya.

Yogyakarta, Mei 2011

(Bernadetha Ayu Setianingtyas)
22074226

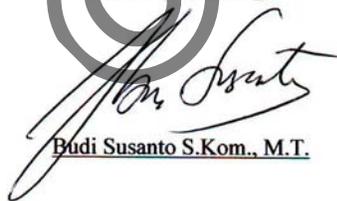
HALAMAN PERSETUJUAN

Judul : **Implementasi Metode Fuzzy K-Nearest Neighbor dalam
Klasifikasi Tugas Akhir Sebagai Penentuan Bidang
Minat**
Studi Kasus Dokumen Abstraksi Tugas Akhir Fakultas
Teknologi Informasi Program Studi Teknik Informatika

Nama : Bernadetha Ayu Setianingtyas
NIM : 22074226
Mata Kuliah : Tugas Akhir
Kode : IN4036
Semester : Genap
Tahun Akademik : 2010/2011

Telah diperiksa dan disetujui
Di Yogyakarta,
Pada Tanggal 28 April 2011

Dosen Pembimbing I



Budi Susanto S.Kom., M.T.

Dosen Pembimbing II



Umi Probayekti S.Kom., MLIS.

HALAMAN PENGESAHAN

SKRIPSI

**IMPLEMENTASI METODE FUZZY K-NEAREST NEIGHBOR
DALAM KLASIFIKASI TUGAS AKHIR SEBAGAI PENENTUAN
BIDANG MINAT**

Studi Kasus Dokumen Abstraksi Tugas Akhir Fakultas Teknologi Informasi
Program Studi Teknik Informatika

Oleh : Bernadetha Ayu Setianingtyas / 22074226

Dipertahankan di depan dewan Penguji Tugas Akhir/Skripsi
Program Studi Teknik Informatikan Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana –Yogyakarta

Dan dinyatakan diterima untuk memenuhi salah satu

Syarat memperoleh gelar

Sarjana Komputer

Pada Tanggal

11 Mei 2011

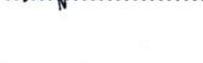
Yogyakarta, 25/5/2011

Mengesahkan,

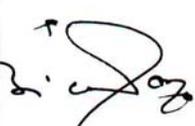
1. Budi Susanto, S.Kom., M.T
2. Umi Proboyekti, S.Kom., MLIS
3. Ir. Sri Suwarno, M.Eng.
4. Drs. Jong Jek Siang, M.Sc.


.....

.....

.....

.....

Dekan



(Drs. Wimmie Handiwidjojo, MIT)

Ketua Program Studi


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



Universitas Kristen Duta Wacana
Fakultas Teknik Program Studi Teknik Informatika
Jl. Dr. Wahidin Sudirohusodo 5 – 25 Yogyakarta 55224
Telp. : (0274) 563929 Faks. : (0274) 513235

FORMULIR PERBAIKAN (REVISI) TUGAS AKHIR

Dengan ini kami menyatakan bahwa mahasiswa yang melakukan Tugas Akhir dibawah ini:

Nama Mahasiswa : Bernadetha Ayu Setianingtyas
NIM : 22074226
Judul Tugas Akhir : **IMPLEMENTASI METODE FUZZY K-NEAREST NEIGHBOR
DALAM KLASIFIKASI TUGAS AKHIR SEBAGAI
PENENTUAN BIDANG MINAT**
Studi Kasus Dokumen Abstraksi Tugas Akhir Fakultas
Teknologi Informasi Program Studi Teknik Informatika

Tanggal Pendaran : 11 Mei 2011

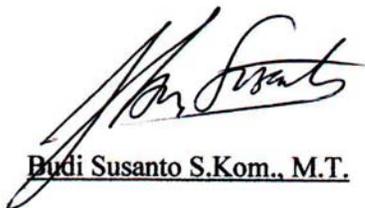
Tanggal Revisi : 21 Mei 2011

Telah melakukan perbaikan tugas akhir dengan lengkap.

Dengan demikian pernyataan kami agar dapat dipergunakan sebagaimana mestinya.

Yogyakarta,

Dosen Pembimbing I


Budi Susanto S.Kom., M.T.

Dosen Pembimbing II


Umi Proboyekti S.Kom., MLIS.

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan anugerah, sehingga penulis dapat menyelesaikan Tugas Akhir dengan baik dan tepat waktu.

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. Bapa Maha Pengasih yang senantiasa memberikan kesempatan untuk terus belajar dan kelancaran dalam semua proses yang telah penulis jalani.
2. **Bpk. Budi Susanto, S.Kom., M.T.** selaku dosen pembimbing I yang telah memberikan bimbingannya dengan sabar dan baik kepada penulis.
3. **Ibu Umi Proboyekti, S.Kom., MLIS** selaku dosen pembimbing II atas bimbingan petunjuk dan masukan yang diberikan selama pengerjaan tugas akhir ini sejak awal hingga akhir.
4. Keluarga tercinta yang memberikan dukungan dan semangat.
5. Orang-orang terdekat yang telah memberikan dukungan dan semangat.
6. Teman-teman yang telah memberikan masukan dan semangat.

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 dalam penyusunan laporan maupun yang pernah penulis lakukan sewaktu membuat program Tugas Akhir. Sekali lagi penulis mohon maaf yang sebesar-besarnya. Dan semoga ini bisa berguna bagi kita semua.

Yogyakarta, Mei 2011

Penulis

© UKDWN

INTISARI

Pencarian topik Tugas Akhir bukan merupakan hal yang mudah, banyak cara yang dilakukan mahasiswa untuk mendapatkannya, salah satunya dengan membaca refesensi dari Tugas Akhir yang pernah dibuat. Repository SInTA menyediakan kumpulan dokumen Tugas Akhir yang pernah disusun oleh mahasiswa UKDW. Namun demikian metadata dalam SInTA tidak menginformasikan mengenai bidang minat pada setiap dokumennya. Sementara itu bidang minat berguna dalam pencarian topik guna menyesuaikan topik tugas akhir dengan kemampuan dan minat mahasiswa.

Metode klasifikasi *K-Nearest Neighbor* telah banyak dikembangkan untuk meningkatkan keakuratan klasifikasi, salah satu pengembangannya adalah metode *Fuzzy K-Nearest Neighbor*. Metode *Fuzzy K-Nearest Neighbor* digunakan dalam penelitian ini untuk melakukan klasifikasi pada dokumen tugas akhir ke dalam kategori bidang minat. Kemampuan klasifikasi metode *Fuzzy K-Nearest Neighbor* dengan hasil probabilitas untuk setiap kategori dan probabilitas terbesar diberikan pada jarak terpendek menghasilkan presisi yang lebih tinggi dari metode *K-Nearest Neighbor*.

Hasil klasifikasi sangat dipengaruhi oleh data *training set* yang digunakan, oleh sebab itu jumlah dan konten dokumen yang akan digunakan sebagai *training set* harus ditentukan dengan benar sehingga dapat mewakili setiap kategori. *Preprocessing* dokumen teks sebelum klasifikasi membutuhkan waktu yang cukup lama oleh karena itu penggunaan *memory* dan *stored database* perlu dikembangkan untuk peningkatan kecepatan eksekusi program.

DAFTAR ISI

HALAMAN JUDUL	
PERNYATAAN KEASLIAN SKRIPSI.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
UCAPAN TERIMA KASIH.....	iv
INTISARI.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Hipotesa	4
1.6 Metode Penelitian	4
1.7 Sistematika Penulisan.....	5
BAB 2 TINJAUAN PUSTAKA.....	6
2.1 Tinjauan Pustaka.....	6
2.2 Landasan Teori.....	7
2.2.1 <i>Text Mining</i>	7
2.2.2 Algoritma Porter untuk Bahasa Indonesia	11
2.2.3 <i>TF – IDF(Term Frequency – Invers Document Frequency)</i>	15
2.2.4 <i>Feature Selection</i>	16
2.2.5 Algoritma <i>K- Nearest Neighbor</i>	17
2.2.6 Algoritma <i>Fuzzy K- Nearest Neighbor</i>	18
2.2.7 Evaluasi Sistem	22
BAB 3 PERANCANGAN SISTEM	23

3.1	Kebutuhan <i>Hardware</i> dan <i>Software</i>	23
3.1.1	Kebutuhan <i>Hardware</i>	23
3.1.2	Kebutuhan <i>Software</i>	23
3.2	Analisis Data	23
3.3	Rancangan Kerja Sistem	26
3.3.1	Spesifikasi Sistem	26
3.3.2	Rancangan Sistem	28
3.4	Block Diagram	28
3.5	Perancangan <i>Database</i>	30
3.5.1	Kamus Data	30
3.5.2	Relasi <i>Database</i>	34
3.6	Rancangan Alur Kinerja Sistem	34
3.6.1	Permintaan Data Tugas Akhir pada <i>web service</i> SInTA	35
3.6.2	<i>Index Construction</i> : Tokenisasi	38
3.6.3	<i>Index Construction</i> : Filterisasi <i>Stopwords</i>	39
3.6.4	<i>Index Construction</i> : <i>Stemming</i>	40
3.6.5	<i>Index Construction</i> : Proses TF-IDF	42
3.6.6	Proses <i>Feature Selection</i>	43
3.6.7	Proses Klasifikasi Tugas Akhir	45
3.7	Rancangan Pengujian Sistem	46
3.8	Rancangan <i>User Interface</i>	47
3.8.1	Rancangan Halaman Input Kata Kunci	47
3.8.2	Rancangan Halaman Daftar Judul	48
3.8.3	Rancangan Halaman Daftar <i>Training Set</i>	49
3.8.4	Rancangan Halaman Hasil Klasifikasi	50
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM.....		52
4.1	Implementasi	52
4.1.1	Antar Muka Sistem.....	52
a.	Halaman Muka	52
b.	Halaman Hasil Pencarian	53
c.	Halaman Klasifikasi	54

d. Halaman Daftar <i>Training Set</i>	55
4.1.2 Implementasi Metode	56
a. Proses Pencarian Judul dan Abstraksi Tugas Akhir	56
b. Proses Pengolahan Data Judul dan Abstraksi	57
c. Proses <i>Preprocessing</i> : Tokenisasi	58
d. Proses <i>Preprocessing</i> : Filterisasi <i>Stopwords</i>	58
e. Proses <i>Preprocessing</i> : <i>Stemming</i>	59
f. Proses <i>Preprocessing</i> : Normalisasi TF-IDF	59
d. Proses <i>Feature Selection</i>	61
d. Proses Klasifikasi	62
4.2 Analisis Sistem	63
4.2.1 Pemilihan <i>Feature Selection</i>	63
4.2.1 Analisa Hasil Klasifikasi K-Nearest Neighbor dan Fuzzy	
K- Nearest Neighbor	63
BAB 5 KESIMPULAN DAN SARAN	73
5.1 Kesimpulan	70
5.2 Saran	70
DAFTAR PUSTAKA	72
LAMPIRAN	



DAFTAR GAMBAR

Gambar 2.1 Proses Algoritma Porter untuk Bahasa Indonesia	12
Gambar 3.1 Block Diagram Sistem	29
Gambar 3.2 Relasi Tabel	34
Gambar 3.3 Flowchart Rancangan Seluruh Kerja Sistem	35
Gambar 3.4 Flowchart Proses Permintaan Data	36
Gambar 3.5 Flowchart Proses Pengolahan Data	37
Gambar 3.6 Flowchart <i>Index Construction</i> Proses Tokenisasi	38
Gambar 3.7 Flowchart <i>Index Construction</i> Proses Filterisasi <i>Stopwords</i>	39
Gambar 3.8 Flowchart <i>Index Construction Stemming</i>	41
Gambar 3.9 Flowchart <i>Index Construction</i> Algoritma Porter	42
Gambar 3.10 Flowchart <i>Index Construction</i> Pembobotan TF-IDF	43
Gambar 3.11 Flowchart <i>Feature Selection</i>	44
Gambar 3.12 Flowchart Klasifikasi Algoritma Fuzzy K-NN	46
Gambar 3.13 Rancangan Halaman Input Kata Kunci	47
Gambar 3.14 Rancangan Halaman Daftar Judul	48
Gambar 3.15 Rancangan Halaman Daftar <i>Training Set</i>	49
Gambar 3.16 Rancangan Halaman Hasil Klasifikasi	50
Gambar 4.1 Halaman Awal	52
Gambar 4.2 Halaman Hasil Pencarian	53
Gambar 4.3 Halaman Klasifikasi	54
Gambar 4.4 Halaman Daftar <i>Training set</i>	56
Gambar 4.5 Klasifikasi dengan Metode <i>K-Nearest Neighbor</i>	65
Gambar 4.6 Frekuensi Kemunculan Hasil Klasifikasi Ganda	65
Gambar 4.7 Klasifikasi dengan Metode <i>K-Nearest Neighbor</i>	66
Gambar 4.8 Grafik Presisi Hasil Klasifikasi Metode <i>K-NN</i> dan <i>Fuzzy K-NN</i> ...	69

DAFTAR TABEL

Tabel 2.1 Daftar Partikel Bahasa Indonesia	13
Tabel 2.2 Daftar <i>Possesive Pronoun</i> Bahasa Indonesia	13
Tabel 2.3 Daftar Aturan Awalan Pertama	14
Tabel 2.4 Daftar Aturan Awalan Kedua	14
Tabel 2.5 Daftar Aturan Akhiran	15
Tabel 2.6 Tabel Vektor Jarak	20
Tabel 3.1 Tabel <i>Training Set</i>	30
Tabel 3.2 Tabel Daftar Judul	30
Tabel 3.3 Tabel Token <i>Training Set</i>	31
Tabel 3.4 Tabel Token Dokumen Uji	31
Tabel 3.5 Tabel Kategori Klasifikasi	32
Tabel 3.6 Tabel <i>Feature Selection</i>	32
Tabel 3.7 Tabel Jarak	33
Tabel 3.8 Tabel <i>Stoplist</i>	33
Tabel 3.9 Tabel Kombinasi	33
Tabel 4.1 Tabel Data Feature Selection	64
Tabel 4.2 Tabel Frekuensi Kemunculan Hasil Klasifikasi Ganda	65
Tabel 4.3 Tabel Hasil Klasifikasi Metode <i>K-NN</i> dan <i>Fuzzy K-NN</i>	67
Tabel 4.4 Tabel Presisi Hasil Klasifikasi Metode <i>K-NN</i> dan <i>Fuzzy K-NN</i>	69

Bab 1

PENDAHULUAN

1.1 Latar Belakang

Skripsi merupakan tugas akhir mahasiswa S1 yang bersifat mandiri dan wajib untuk mendapatkan gelar sarjana. Seorang mahasiswa yang akan menulis tugas akhir harus mencari topik penelitian terlebih dahulu untuk menentukan fokus penelitiannya. Ada banyak cara yang dilakukan mahasiswa dalam usaha pencarian topik tugas akhir, seperti membaca jurnal penelitian dalam *e-journal*, mengikuti penelitian yang dilakukan dosen, membaca kumpulan tugas akhir yang pernah dibuat, melakukan observasi masalah yang sering timbul di bidangnya dan sebagainya. Banyak cara mendapatkan topik tugas akhir, akan tetapi mencari topik tugas akhir bukanlah hal yang mudah, hal ini terbukti dengan banyak mahasiswa yang kesulitan untuk memulai mengerjakan tugas akhir karena belum mendapatkan topik yang sesuai.

Saat ini telah banyak universitas ternama yang memiliki *repository* tugas akhir berbasis web yang dapat dikunjungi oleh pihak di luar universitas. Layanan ini dapat dimanfaatkan oleh mahasiswa untuk mendapatkan ide topik tugas akhir. Akan tetapi data tugas akhir yang tersimpan dalam *repository* digital tersebut dirasa masih memiliki kekurangan. Sebagai contoh ketika seorang mahasiswa Teknik Informatika membaca sebuah tugas akhir mengenai metode *Forward Chaining*. Mahasiswa yang belum pernah mendengar mengenai metode ini tentunya akan memikirkan matakuliah apa yang membahas mengenai metode *Forward Chaining* atau bidang minat apa yang berkaitan metode tersebut.

Universitas Kristen Duta Wacana telah memiliki SInTA sebagai *repository* tugas akhir berbasis web, menyediakan layanan informasi data tugas akhir dengan menampilkan judul dan abstraksi. Layanan ini cukup membantu mahasiswa UKDW dan mahasiswa umum untuk melakukan pencarian topik

tugas akhir. Akan tetapi, data tugas akhir yang dimiliki oleh SInTA masih memiliki kekurangan mengenai metadata bidang minat dan matakuliah yang dapat membantu pencarian topik tugas akhir lebih spesifik.

Penelitian ini akan difokuskan pada penentuan bidang minat dari tugas akhir melalui informasi dalam dokumen abstraksi dan judul tugas akhir, dengan asumsi abstraksi dan judul mewakili isi penulisan tugas akhir. Penelitian dilakukan dengan mengklasifikasi data dokumen teks abstraksi dan judul ke kategori bidang minat yang ada dengan menggunakan salah satu metode klasifikasi dalam *Text Mining* yaitu *Fuzzy K-Nearest Neighbor*.

K-Nearest Neighbor merupakan salah satu metode klasifikasi yang telah lama digunakan untuk klasifikasi dokumen teks. Metode ini telah banyak mengalami perkembangan guna mendapat nilai presisi klasifikasi yang lebih baik. Salah satunya *Fuzzy K-Nearest Neighbor*, metode yang menggabungkan metode *K-Nearest Neighbor* dengan metode *fuzzy* untuk mengukur ketepatan hasil klasifikasi yang dilakukan. Metode ini akan menghasilkan nilai keanggotaan bagi setiap kelas kategori. Dengan demikian, penelitian akan mendapatkan nilai presisi klasifikasi yang lebih baik dan sebaran nilai keanggotaan untuk masing-masing kategori.

1.2 Perumusan Masalah

Seberapa tinggi presisi sistem klasifikasi yang didapat dengan menggunakan Algoritma *Fuzzy K-Nearest Neighbor*. Penghitungan presisi dari ketepatan klasifikasi dokumen tugas akhir yang diwakili oleh abstraksi dan judul.

1.3 Batasan Masalah

Batasan-batasan masalah dari penelitian ini:

1. Klasifikasi akan dilakukan pada file teks Abstraksi Tugas Akhir dalam database SInTA dengan memanfaatkan layanan *Web Service*.

2. Studi kasus yang diambil pada Bidang Minat Fakultas Teknologi Informasi Prodi Teknik Informatika.
3. Pembentukan vektor dokumen diperoleh dari abstraksi dan judul, dengan asumsi judul dari tugas akhir mewakili keseluruhan isi penulisan sehingga pemberian bobot pada judul akan dibuat lebih besar dengan menambah frekuensi kemunculan tokennya.
4. Besarnya prosentase nilai *feature selection* dan *k* ditentukan oleh masukan dari pengguna.
5. Klasifikasi dilakukan dengan menggunakan algoritma *Fuzzy K-Nearest Neighbor* dan *K-Nearest Neighbor* untuk mengetahui perbedaan presisi yang dihasilkan.

1.4 Hipotesa

Penggunaan algoritma *Fuzzy K-Nearest Neighbor* dan peningkatan bobot token judul dengan asumsi bahwa judul tugas akhir mewakili keseluruhan isi penulisan pada klasifikasi dokumen tugas akhir, dapat menghasilkan nilai presisi yang lebih baik. Nilai *fuzzy* yang dihasilkan dari proses klasifikasi menunjukkan nilai keanggotaan terhadap masing-masing kelas kategori.

1.5 Tujuan Penelitian

Tujuan dari penelitian ini antara lain,

1. Mengaplikasikan teori *Text Mining* dengan mengklasifikasi dokumen tugas akhir ke dalam Bidang Minat dengan mengambil studi kasus untuk Program Studi Teknik Informatika.
2. Mengaplikasikan proses *feature selection* untuk membantu kinerja proses klasifikasi.
3. Mengaplikasikan algoritma *Fuzzy K-Nearest Neighbor* untuk mendapatkan hasil presisi yang lebih baik dan membantu mengurangi terjadinya hasil klasifikasi ganda atau kesamaan nilai klasifikasi yang sering terjadi pada algoritma *K-Nearest Neighbor*.

1.6 Metode Penelitian

Metode yang digunakan melakukan penelitian ini adalah algoritma *Fuzzy K-NN* (*K-Nearest Neighbor*) dan metode yang digunakan untuk mengumpulkan data dan sumber informasi melalui studi pustaka dalam penelitian ini dengan menitik beratkan pada:

- a) Mempelajari proses pengambilan informasi melalui layanan *web service*, dalam hal ini juga bertujuan untuk mengembangkan modul dalam pengaksesan database tugas akhir melalui *web service* SinTA.
- b) Melakukan penghitungan vektor dokumen dengan memperhatikan penambahan bobot token pada judul dan proses stemming pada penerapan dalam teks berbahasa Indonesia.
- c) Mempelajari pemberian *feature selection* yang tepat untuk mendapatkan nilai presisi yang baik.
- d) Melakukan proses klasifikasi pada dokumen teks dan penerapan algoritma *Fuzzy K-NN* pada proses klasifikasi dokumen teks, sekaligus pengembangan modul *K-Nearest Neighbor* dalam proses klasifikasi guna mendapatkan nilai presisi yang lebih baik.
- e) Melakukan pengujian terhadap keseluruhan proses yang terjadi dalam sistem guna mendapatkan hasil klasifikasi yang diinginkan.

1.7 Sistematika Penulisan

Sistematika penulisan karya tulis ini akan terbagi dalam 5 bab yang akan dibahas sebagai berikut :

Bab I merupakan pendahuluan, yang berisi mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, hipotesis penelitian, metode penelitian dan sistematika penulisan.

Bab II merupakan landasan teori karya tulis, yang berisi mengenai teori-teori klasifikasi dokumen, algoritma fuzzy k-nn serta teori yang mendukung pembangunan sistem. Keseluruhan teori tersebut berasal dari studi pustaka yang dilakukan oleh penulis.

Bab III merupakan desain dan perancangan sistem, yang berisi mengenai perancangan sistem dan desain sistem seperti flowchart program, desain database, *user interface*, spesifikasi sistem serta rancangan pengujian sistem.

Bab IV merupakan implementasi dan analisis sistem, yang menerangkan mengenai sistem yang telah dibangun, pengujian sistem serta hasil pengujian sistem yang telah dilakukan. Hasil pengujian tersebut akan dicocokkan kembali dengan tujuan awal penelitian sesuai yang tertulis di Bab I.

Bab V merupakan kesimpulan dan saran, yang berisi dari hasil penelitian yang telah dilakukan dan evaluasi sistem yang membahas kelebihan dan kekurangan sistem. Hasil evaluasi tersebut dapat menjadi saran untuk pengembangan penelitian dengan topik yang sama kedepannya.

© UKDW

Bab 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan evaluasi dan analisa hasil klasifikasi, diperoleh kesimpulan sebagai berikut,

1. Penerapan *stemming* pada dokumen teks berbahasa Indonesia perlu diperhatikan lebih, dengan adanya pemakaian kata asing dalam dokumen yang dapat kehilangan makna sebenarnya karena proses *stemming*.
2. Metode *Fuzzy K - Nearest Neighbor* memberikan kontribusi besar pada kategori yang memiliki jarak terpendek terhadap dokumen uji untuk menghasilkan nilai probabilitas tinggi.
3. Metode *Fuzzy K - Nearest Neighbor* mengurangi resiko terjadinya hasil klasifikasi ganda pada metode *K - Nearest Neighbor* karena menghasilkan nilai probabilitas pada masing-masing kelas kategori. Klasifikasi menggunakan metode *Fuzzy K - Nearest Neighbor* memberi nilai probabilitas tinggi pada vektor jarak dengan nilai terpendek. Nilai presisi yang dihasilkan oleh metode *Fuzzy K - Nearest Neighbor* lebih tinggi dari metode *K - Nearest Neighbor*.

5.2. Saran

Saran yang diajukan penulis untuk pengembangan dan perbaikan sistem untuk hasil yang lebih optimal adalah :

1. Proses *stemming* teks berbahasa Indonesia dengan memanfaatkan penyimpanan kata dasar dalam kamus akan menghasilkan kata dasar yang lebih baik bila dilakukan penambahan kata-kata asing baik yang berupa kata serapan atau tidak. Hal ini dilakukan untuk mengurangi kesalahan proses *stemming* yang dapat mengakibatkan hilangnya kata asing yang memiliki informasi penting dalam dokumen teks.

2. Pemakaian *database stoplist*, kamus dan kata berimbuhan dapat di optimalkan dengan menambahkan kemampuan *update* data pada sistem.
3. Fungsi pengecekan frase dalam dokumen teks perlu ditambahkan guna mencegah hilangnya frase penting dalam proses tokenisasi.
4. Jenis topik tugas akhir dalam penelitian ini sangat luas pada masing-masing kategori, dengan demikian dapat dilakukan penambahan jumlah data *training set* guna menaikkan nilai presisi klasifikasi.
5. Jumlah dokumen yang besar menghasilkan himpunan vektor yang cukup besar dan mengakibatkan proses *index construction* semakin lama terlebih dalam proses update database. Perbaikan sistem dalam hal akses database dan penghitungan nilai dengan memanfaatkan penggunaan *memory* dapat mempercepat waktu proses.



DAFTAR PUSTAKA

- Arif, Muhammad, Muhammad Usman Akram dan Fayyaz-ul-Afsar Amir Minhas. (2010). Pruned fuzzy K-nearest neighbor classifier for Beat. *J.Biomedical Science and Engineering*, vol(3), hal 380-389.
- Han, Jiawei & Micheline Kamber. (2006). *Data Mining: Concepts and Technique 2nd Edition*. San Fransisco : Morgan Kaufmann Publishers.
- Hs, Widjono.(2007). *Bahasa Indonesia Mata Kuliah Pengembangan Kepribadian di Perguruan Tinggi*. Jakarta : Grasindo.
- Huang, Ying & Yanda Li. (2004). *Prediction of Protein subcellular locations using fuzzy k-NN method*. Oxford University Press. Diakses terakhir kali pada tanggal 28 April 2011 dari <http://bioinformatics.oxfordjournals.org/content/20/1/21.full.pdf#page=1&view=FitH>
- Keller, J.M., Gray,M.R. & Given, J.A.(1985). A Fuzzy K-Nearest Neighbour Algorithm. *IEEE Transaction IEEE Transactions on Systems, Man, and Cybernetics*, vol(15) 580-585. Diakses terakhir pada tanggal 28 April 2011 dari <http://www.cs.missouri.edu/~skubicm/8820/FuzzyKNN.pdf>
- Manning, Christopher D., Raghavan, P., & Schutze, H. (2008). *Introduction to Information Retrieval*. United State of America : Cambrige University Press .
- Moeliono, Anton. M.(1993). *Tata Bahasa Baku Bahasa Indonesia*. Jakarta : Balai Pustaka.

Porter, M.(1980) *An algorithm for suffix stripping*. Program 13(3), 130-137.
Diakses terakhir pada tanggal 28 April 2011 dari
<http://www.mis.yuntech.edu.tw/~huangcm/research/porter.pdf>

Tala, Fadilla Z.(2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. University of Amsterdam. Diakses terakhir kali pada 28 April 2011 dari <http://www.ilc.uva.nl/Publications/ResearchReports/MoL-2003-02.text.pdf>

Weiss, Sholom M., Nitin Indurkha, Tong Zhang dan Fred J. Damerau. (2005). *Text Mining*. United States of America : Springer.

