Implementasi Algoritma *Fuzzy* K-*Nearest Neighbour* (*Fuzzy* K-NN) Untuk Klasifikasi Proposal Skripsi Berdasarkan Kelompok Keilmuan Di Teknik Informatika UNIKOM

Yosep Amin

Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
Email : yosep.amin@gmail.com

ABSTRAK

Skripsi merupakan mata kuliah penutup yang diambil oleh mahasiswa/mahasiswi yang sedang menempuh jenjang pendidikan strata 1 di perguruan tinggi negeri atau pun perguruan tinggi swasta, skripsi menjadi salah satu syarat untuk kelulusannya. Berkembangnya ilmu pengetahuan setiap waktunya membuat semakin beragamnya tema – tema skripsi yang bermunculan, sehingga menimbulkan adanya keanekaragaman proposal skripsi. Sehingga pihak akademik membuat sebuah pengelompokan kepada tema – tema skripsi tersebut Walaupun demikian masih saja terjadi kesalahan pengelompokan proposal skripsi. Faktor pemicu salah memilih kelompok keilmuan karena masih berkaitannya satu kelompok keilmuan dengan kelompok keilmuan yang lain. Untuk melakukan identifikasi proposal skripsi dapat dilakukan dengan memanfaatkan metode klasifikasi.
Dalam penelitian ini metode *fuzzy* k-*nearest neighbour* (FK-NN) diimplementasikan untuk klasifikasi proposal skripsi berdasarkan kelompok keilmuan. Dengan melakukan pengolahan kata sebelumnya serta menggunakan TFIDF sebagai metode pembobotan kata sehingga kata input dapat diolah sesuai kebutuhan. Parameter yang digunakan sebagai faktor penilaian yakni dilihat dari judul dan deskripsi proposal skripsi.

Berdasarkan hasil pengujian didapatkan niai keakuratan yang terbesar yakni dengan presentase sebesar 51% pada nilai K=7 dan K=8, maka dari itu dapat disimpulkan bahwa dalam penelitian yang dilakukan nilai K yang menghasilkan keakuratan terbaik yakni pada nilai K=7 dan K=8.

Kata kunci : proposal skripsi, *Fuzzy* K-*Nearest Neighbour*, TF-IDF, Klasifikasi.

1. PENDAHULUAN

1.1 Latar Belakang

Skripsi merupakan matakuliah penutup yang diambil oleh mahasiswa/mahasiswi yang sedang menempuh jenjang pendidikan strata 1 di perguruan tinggi negeri atau pun perguruan tinggi swasta, skripsi menjadi salah satu syarat untuk kelulusannya. Skripsi merupakan istilah yang digunakan di indonesia untuk mengilustrasikan suatu karya tulis ilmiah berupa paparan tulisan hasil penelitian sarjana S1 yang membahas permasalahan/fenomena dalam bidang ilmu tertentu dengan menggunakan kaidah–kaidah yang berlaku[1]. Karya tulis/penelitian ilmiah ini akan dilaporkan kedalam laporan skripsi, sebelum melakukan skripsi, biasanya para mahasiswa/mahasiswi akan membuat terlebih dahulu usulan penelitian atau biasanya disebut proposal skripsi yang akan diajukan ke panitia sidang proposal skripsi untuk disidangkan.

Berkembangnya ilmu pengetahuan setiap waktunya membuat semakin beragamnya tema – tema skripsi yang dibermunculan yang diajukan sebagai tema penelitian, yang kemudian dituangkan dalam proposal skripsi, dari perbedaan tema proposal skripsi yang dibuat oleh mahasiswa, menimbulkan adanya keanekaragaman proposal skripsi baik dari segi latar belakang, tujuan, metodologi, atau data penelitian. Sehingga pihak akademik membuat sebuah pengelompokan kepada tema – tema skripsi tersebut, ini dilakukan agar proposal skripsi dapat dikenali dan membantu menempatkan proposal skripsi ke kelompok keilmuan yang sesuai. Walaupun demikian masih saja terjadi kesalahan pengelompokan proposal skripsi, mahasiswa terkadang salah memilih kelompok keilmuan. Jika ini terjadi maka penanganan terhadap proposal skripsi seperti masukan untuk arah pengerjaan skripsi tersebut akan tidak maksimal, karena proposal skripsi tidak ditangani oleh penguji yang ahli di bidangnya. Faktor pemicu salah memilih kelompok keilmuan karena masih berkaitannya satu kelompok keilmuan dengan kelompok keilmuan yang lain. Untuk melakukan identifikasi proposal skripsi dapat dilakukan dengan memanfaatkan metode klasifikasi.

Metode *Fuzzy k-Nearest Neighbour* merupakan salah satu metode yang digunakan untuk klasifikasi, metode ini dikembangkan pertama kali oleh James M. Keller dasar dari algoritma ini adalah pemberian nilai *membership* sebagai fungsi nilai pola jarak dari sejumlah himpunan *k-Nearest Neighbour* dan pemberian nilai keanggotaan *Neighbour* pada kelas

tertentu[2]. Pada penelitian yang pernah dilakukan yang berhubungan dengan kasus ini yaitu berjudul *Web Document Classification Based on Fuzzy k-NN Algorithm* yang dilakukan oleh Zhang, Yi Niu, dan Huabei nie[2]. Mengatakan kelebihan dari metode *fuzzy* k-NN terletak pada tingkat keakurasian yang dihasilkan lebih baik daripada metode klasifikasi lain seperti k-*Nearest Neighbour* (k-NN), dan *Support Vektor Machine*(SVM). Pada penelitian tersebut juga mengatakan bahwa kekurangan yang dimiliki oleh *Fuzzy* k-NN ini, yakni terletak pada kecepatan pengklasifikasian yang lebih lambat dibandingkan dua metode klasifikasi pembanding diatas[2].

Berdasarkan penjelasan diatas, dalam penelitian ini akan dilakukan implementasi algoritma fuzzy k-*Nearest Neighbour* untuk menyelesaikan permasalahan kasus klasifikasi proposal skripsi. Diambil studi kasus di program studi Teknik Informatika UNIKOM.

1.2 Rumusan Masalah

Berdasarkan pemaparan latar belakang diatas dapat dirumuskan masalahnya, antara lain : Bagaimana implementasi metode *Fuzzy k-Nearest Neighbour* untuk melakukan klasifikasi proposal skripsi.

1.3 Maksud

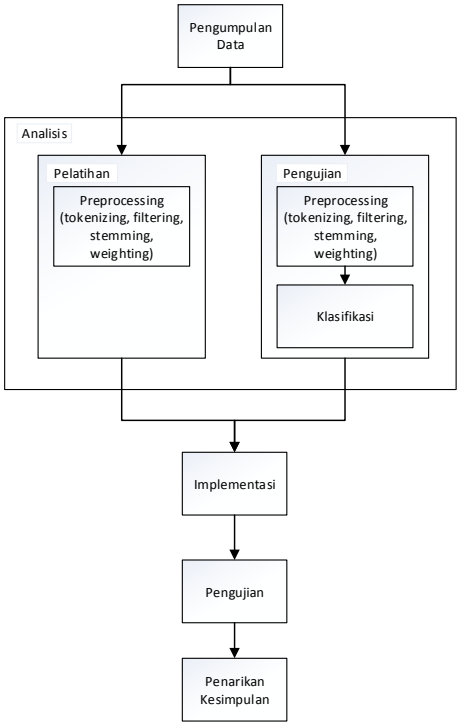
Maksud dari penelitian ini adalah untuk implementasi *Fuzzy k-Nearest Neighbour* dalam melakukan klasifikasi Proposal Skripsi.

1.4 Tujuan

Tujuan dari penelitian ini adalah untuk mengetahui keakuratan metode *Fuzzy k-Nearest Neighbour* dalam melakukan klasifikasi proposal skripsi.

1.5 Metodologi Penelitian

Metode penelitian yang digunakan dalam penelitian ini yaitu metode penelitian deskriptif. Merupakan suatu metode dalam meneliti status sekelopok manusia, suatu objek, suatu set kondisi, suatu sistem pemikiran, ataupun suatu kelas peristiwa pada masa sekarang. Tujuan dari penelitian deskriptif ini adalah untuk membuat deskripsi, gambaran atau lukisan secara sistematis, faktual dan akurat mengenai fakta-fakta, sifat-sifat serta hubungan antar fenomena yang diselidiki. Serta memiliki ciri-ciri menerangkan hubungan, menguji hipotesis-hipotesis, membuat prediksi serta mendapatkan makna(kesimpulan) dan implikasi dari suatu masalah yang ingin dipecahkan[3]. alur penelitan yang akan dilakukan pada penelitian ini dapat dilihat pada gambar 1.1 sebagai berikut:



Gambar 1 Alur Penelitian

2. ISI PENELITIAN

2.1 Landasan Teori

2.1.1 Proposal Skripsi

Skripsi adalah istilah yang digunakan di Indonesia untuk mengilustrasikan suatu karya tulis ilmiah berupa paparan tulisan hasil penelitian sarjana S1 yang membahas suatu permasalahan/fenomena dalam bidang ilmu tertentu dengan menggunakan kaidah-kaidah yang berlaku[1]. Proposal skripsi sendiri merupakan Suatu permintaan, pengajuan ide kepada pihak lain untuk mendapatkan perizinan topik, tempat penelitian, dan lain sebagainya yang diajukan oleh seorang mahasiswa dalam usahanya untuk melakukan skripsi sebagai syarat kelulusannya sebagai sarjana.

2.1.2 *Text Mining*

Text Mining merupakan penerapan konsep dan teknik data mining untuk mencari pola dalam teks, proses penganalisisan teks guna menyarikan informasi yang bermanfaat untuk tujuan tertentu. Proses data mining untuk data dokumen atau teks memerlukan lebih banyak tahapan, mengingat data teks memiliki karakteristik yang lebih kompleks dari pada data biasa. Tujuannya yakni mencari kata – kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen diproses[7]. Langkah – langkah yang dilakukan pada text mining pada penelitian ini antara lain text preprocessing, term weighting, dan klasifikasi.

2.1.2.1 *Text Preprocessing*

Pada Tahap *text preprocessing* dilakukan beberapa proses menyiapkan judul dan deksripsi proposal skripsi agar menjadi dokumen teks yang siap diolah ditahap selanjutnya, umumnya pada tahap ini terdapat beberapa proses , diantaranya: *tokenizing, filtering, stemming.*

2.1.2.2 *Tokenizing*

Tahap *tokenizing* adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Setiap kalimat yang ada di dokumen akan disusun menjadi kata per kata, semua tanda baca dan tanda hubung akan dihapuskan.

2.1.2.3 *Filtering*

Tahap *Filtering* adalah tahap mengambil kata – kata penting dari hasil token, dalam proses ini bisa menggunakan algoritma *stop list* (membuang kata yang kurang penting) atau *word list*(menyimpan kata penting). *Stop list* adalah daftar kata yang sering digunakan dan tidak menjelaskan isi dari dokumen atau *stopword.*

2.1.2.4 *Stemming*

Tahap *Stemming* adalah tahap dimana kata hasil filtering akan dirubah kedalam bentuk dasarnya(*Stem*). Untuk kata dalam bahasa inggris, metode *stemming* yang paling populer adalah menggunakan algoritma Martin porter's *stammer* atau biasa disebut *porter stemmer.* Algoritma ini juga bisa diaplikasikan untuk bahasa selain inggris, dalam hal ini akan diterapkan pada bahasa indonesia juga.

Porter stemmer for Bahasa Indonesia dikembangkan oleh Fasillah Z. Tala pada tahun 2003. Implementasi *porter stemmer for* Bahasa Indonesia berdasarkan *English Porter Stemmer* yang dikembangkan oleh W.B Frake pada tahun 1992[8].

2.1.2.5 *Term Weighting*

Tahap *Term weighting* adalah tahap pembobotan kata yang dihasilkan dari proses *stemming.* Metode yang sering digunakan yaitu menggunakan metode TF-IDF (*Term Frequency – Inverse Document Frequency*). Pada metode ini, perhitungan bobot *term*(kata) dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Inverse Document Frequency*.

Term Frequency (TF) adalah faktor yang menentukan bobot *term* pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Nilai jumlah kemunculan suatu kata (*term frequency*) diperhitungkan dalam pemberian bobot terhadap suatu kata. Semakin besar jumlah kemunculan suatu term dalam dokumen, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuian yang semakin besar.

Inverse Document Frequency (IDF) adalah pengurangan dominansi *term* yang sering muncul di berbagai dokumen. Hal ini diperlukan karena term yang banyak muncul di berbagai dokumen, dapat dianggap sebagai term umum (*common term*) sehingga tidak penting nilainya. Sebaliknya faktor kejarangmunculan kata (*term scarcity*) dalam koleksi dokumen harus diperhatikan dalam pemberian bobot. Kata yang muncul pada sedikit dokumen harus dipandang sebagai kata yang lebih penting (*uncommon term*) daripada kata yang muncul pada banyak dokumen. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu kata (*inverse document frequency*). Hal ini merupakan usulan dari George Zipf. Zipf mengamati bahwa frekuensi dari sesuatu cenderung kebalikan secara proposional dengan urutannya[2].

Untuk menghitung bobot kata (Wi) dalam dokumen dihitung dengan menggunakan rumus:

$$w_i = TF(t_i, d) \times IDF(t_i) \qquad (1)$$

Dimana:

w_i = bobot kata *(term)* dalam dokumen d.
i = indeks kata (*term).*
$TF(t_i, d)$ = banyaknya kata (*term*) t_i yang muncul dalam dokumen d
$IDF(t_i)$ = *inverse document frequency* dari kata (*term*) t_i.

Untuk menghitung $IDF(t_i)$ dapat menggunakan rumus:

$$IDF(t_i) = \log\left(\frac{D}{DF(t_i)}\right) \qquad (2)$$

Dimana:

$IDF(t_i)$ = *inverse document frequency* dari kata (*term*) t_i.
D = jumlah dokumen keseluruhan.
$DF(t_i)$ = jumlah dokumen yang memiliki kata (*term*) (t_i).

2.1.2.6 *Fuzzy* K-*Nearest Neighbour*

Algoritma ini memberikan nilai membership sebagai fungsi pola jarak dari sejumlah himpunan K-*nearest neighbour* dan pemberian nilai keanggotaan *neighbour* pada kelas tertentu. Sehingga pada algoritma ini data *testing* yang akan diklasifikasikan akan memiliki nilai keanggotaan pada semua kelas. Klasifikasi algoritma ini nantinya akan memilih nilai keanggotaan kelas pada data *testing* yang paling tinggi[2].

Berikut ini merupakan rumus untuk menghitung jarak antara dokumen testing dan dokumen training dengan menggunakan metode *euclidean distance*:

$$\|x - x_j\| = \left(\sum_{i=1}^{n} |N_i - N_j^i|^2\right)^{1/2} \quad (3)$$

Dimana :

$\|x - x_j\|$= jarak e*uclidean* vektor data *testing* dan dokumen *training* k-*j*.

Ni = bobot *term* dalam dokumen *testing*.

N_i^j = bobot *term* pada dokumen *training* ke-*j*.

i = index *term* ke-*i.*

n = jumlah *term* keseluruhan hasil *text preprocessing*.

Untuk menghitung nilai keanggotaan data testing pada setiap kelas, sebagai berikut:

$$\mu(x) = \frac{\sum_{i=1}^{k} \mu_{ij} \left(\frac{1}{\|x - x_j\|^{2/(m-1)}}\right)}{\sum_{i=1}^{k} \left(\frac{1}{\|x - x_j\|^{2/(m-1)}}\right)} \quad (4)$$

Dimana:

$\mu(x)$ = nilai keanggotaan kelas ke-*i* pada data *testing x*.

M = bobot pangkat (*weight exponent*).

μij = nilai keanggotaan kelas ke-*i* pada tetangga ke-*j*.

Untuk menghitung nilai keanggotaan kelas ke-*i* pada tetangga ke-*j* dihitung menggunakan rumus dibawah ini:

$$\mu ij = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \quad (5)$$

Dimana :

μij = nilai keanggotaan kelas ke-*i* pada tetangga ke-*j*.

I = kelas *genre*.

J = Tetangga terdekat (K).

2.1.2.7 *Confusion Matrix*

Confusion matrix merupakan sebuah metode perhitungan yang digunakan untuk mencari keakuratan pada hasil klasifikasi. Berikut merupakan tabel *confusion matrix*:

Tabel 1 *Confusion Matrix*

Kelas		Prediksi	
		1	0
Kelas sebenarnya	1	TP	FN
	0	FP	TN

Keterangan:

a. *True Positive* (TP), merupakan jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1.
b. *False Positive* (FP), merupakan jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.
c. *False Negative* (FN), merupakan jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.
d. *True Negative* (TN), merupakan jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0. [12]

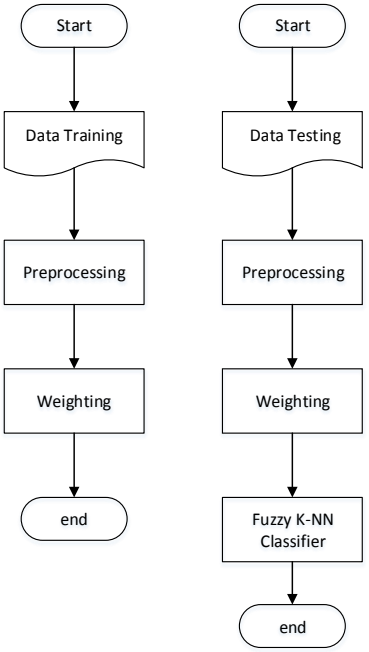
Kemudian akan dihitung akurasinya, akurasi merupakan perhitungan antara jumlah prediksi benar dan dibagikan dengan jumlah prediksi yang dilakukan yang mempunyai tujuan untuk mengetahui berapa presentase keberhasilan suatu sistem melakukan klasifikasi, untuk menghitung akurasi digunakan rumus sebagai berikut:

$$akurasi = = \frac{jumlah\ prediksi\ benar}{jumlah\ seluruh\ prediksi\ yang\ dilakukan} \times 100\ \% \quad (6)$$

2.2 Analisis dan Perancangan Sistem

2.2.1 Analisis Proses

Pada analsis proses akan dijelas proses – proses apa saja yang ada yang digunakan untuk mengolah kata dan sehingga nantinya dapat digunakan untuk proses klasifikasi. Secara umum proses dapat dilihat pada gambar berikut:

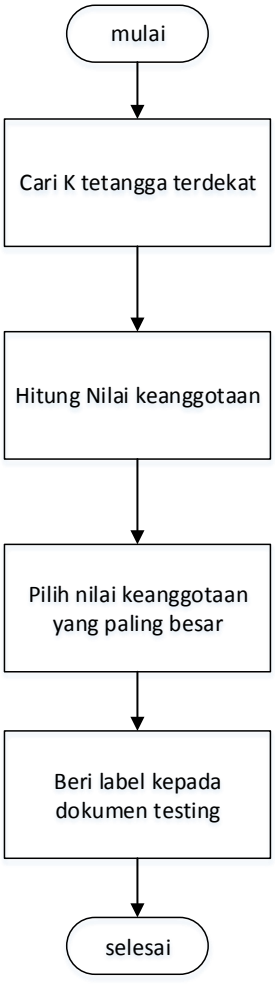


Gambar 2 Analisis Alur Proses

2.2.2 Analisis Proses Klasifikasi

Pada tahap ini akan ditentukan seberapa dekat keterhubungan antara kata – kata pada data yang ada yakni data training sebagai data awal yang berfungsi sebagai informasi dan data testing yang berfungsi sebagai data yang akan dicari kesesuaiannya terhadap data latih. Dalam penelitian ini digunakan algoritma fuzzy k-nearest neighbour (FK-NN). Fknn ini mengkombinasikan kemampuan fuzzy dan algoritma k-nearest neighbour. Kemampuan fuzzy yang diambil yaitu fungsi keanggotaan, dengan fungsi keanggotaan ini akan memetakan sebuah elemen x pada semesta nilai keanggotaan menggunakan sebuah bentuk fungsi teoritis, sehingga dapat diperoleh kemungkinan nilai keanggotaan yang lebih mendekati dann sebagai fungsi pola jarak / kesamaan, dan algoritma k-nearest neighbour akan berperan untuk mengukur seberapa dekat jarak dari nilai keanggotaan yang ada dana akan memilih nilai keanggotaan pada data testing yang paling tinggi.

Untuk lebih jelasnya, langkah – langkah metode *fuzzy* k-*nearest neighbour* adalah sebagai berikut:



Gambar 3 Alur FKNN

2.2.3 Analisis Data

Analisis data terdiri dari analisis data masukan dan analisis data keluaran, sampel data berupa dokumen judul dan deskripsi proposal skripsi telah disediakan sebelumnya.

2.2.3.1 Analisis Data Masukan

Data masukan berupa dokumen yang berisi judul dan deskripsi dari proposal skripsi. Kemudian akan dilakukan proses tokenizing untuk memecah string yang ada didokumen menjadi kata – kata pembentuknya dan menghilangkan tanda baca, kemudian hasilnya akan masuk ke proses selanjutnya yakni proses filtering yaitu proses dimana kata – kata yang kurang penting akan dihapus, kemudian setelah proses filtering, proses yang akan dilakukan selanjutnya yaitu proses stemming, tujuan proses ini yaitu untuk mendapatkan bentuk dasar atau kata dasar dari hasil filtering. Setelahnya akan masuk ketahap selanjutnya. yakni tahap term weighting, ditahap ini akan dilakukan pembobotan kata. Hal ini dilakukan sebagai bekal proses klasifikasi.

2.2.3.2 Analisis Data Keluaran

Data keluaran berupa nama klasifikasi dari dokumen proposal skripsi, data ini didapatkan dari serangkaian proses yang sudah dilakukan sebelumnya, mulai dari tahap preprocessing, tahap term weighting, pelatihan dan, pengujian. Sehingga dari proses – proses yang dilakukan klasifikasi dengan fuzzy k-neareast neigbor dapat dilakukan kemudian dihasilkan dokumen yang terklasifikasi.

2.2.3 Analisis Metode

Analisis metode merupakan penjelasan yang akan dilakukan, berupa tahapan demi tahapan yang ada untuk menyelesaikan persoalan klasifikasi dokumen proposal skripsi ini, analisis algoritma yang akan diteliti pada klasifikasi proposal skripsi berdasarkan judul dan deskripsi ini adalah untuk menganalisis cara kerja *fuzzy* k-*nearest neighbour* dalam mengklasifiksi dokumen proposal skripsi berdasarkan judul dan deskripsi. Sebelum dapat mengklasifikasi, dokumen masukan harus melalui beberapa tahap, yaitu tahap pelatihan, dan tahap pengujian.

2.2.3.1 Analisis Tahap Pelatihan

Tahap pelatihan dilakukan untuk mendapatkan data pembelajaran yang berguna untuk data pembandingan pada proses klasifikasi di tahap pengujian. Tahap pelatihan terdiri dari tahap *preprocessing*, tahap *term weighting*.

2.2.3.2 Analisis Tahap Pengujian

Pada tahap pengujian. Langkah - langkah yang dilakukan untuk tahap pengolahan kata sama seperti pada tahap pelatihan yakni meliputi proses

text preprocessing(*tokenizing*, *filtering*, *stemming*) dan *weighting*. Dokumen testing yang akan diklasifikasi harus melewati proses tersebut. Setelah proses *text preprocessing* dan *weighting* sudah dilakukan, selanjutnya akan dilakukan proses klasifikasi. Langkah- langkahnya adalah sebagai berikut:

1. Cari K tetangga terdekat
   a. Hitung jarak dokumen testing dengan setiap dokumen training, menggunakan rumus 3.
   b. Urutkan hasilnya secara *ascending*
   c. ambil data jarak dokumen sesuai jumlah K.
2. Perhitungan Nilai keanggotaan
   a. Hitung nilai keanggotaan menggunakan rumus 4.
3. Pilih Nilai Keanggotaan Terbesar
4. Beri Label pada dokumen t*esting*.

2.2.4 Analisis Kebutuhan Non fungsional

Kebutuhan non fungsional yang dibutuhkan untuk implementasi algoritma *fuzzy* k-*nearest neighbour* (FK-NN) untuk klasifikasi proposal skripsi kedalam sistem meliputi dua hal yakni kebutuhan perangkat keras dan perangkat lunak.

a. Kebutuhan Perangkat KerasSpesifikai perangkat keras yang digunakan untuk membangun sistem ini diantaranya adalah sebagai berikut:
   1. Prosesor Intel Core i3 2.2 Ghz
   2. Ram 2 GB
   3. Harddisk berkapasitas 500 GB
   4. VGA Nvidia GeForce berkapasitas 1 GB

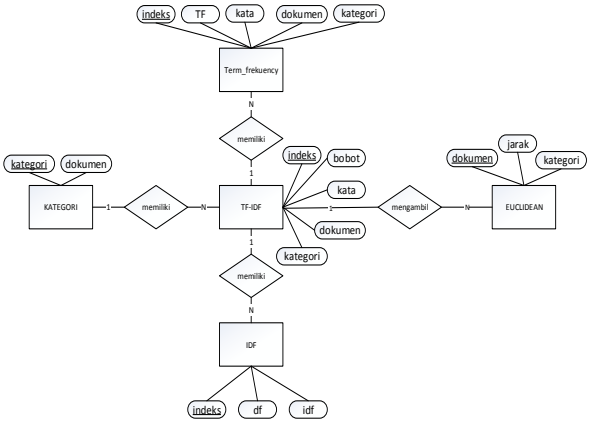
b. Kebutuhan Perangkat Lunak
   Perangkat lunak yang digunakan dalam membangun sistem ini diantaranya adalah sebagai berikut:
   1. Sistem operasi windows 7 home premium
   2. Visual studio 2010

2.2.4 Analisis Kebutuhan Fungsional

Analisis kebutuhan fungsional adalah langkah – langkah mengidentifikasi proses yang akan berjalan didalam sistem yang akan dikembangkan dan menjelaskan kebutuhan yang diperlukan. Sehingga sistem dapat berjalan seperti yang diharapkan. Tahapan pemodelan dalam analisis tersebut sebagai berikut:

2.2.4.1 *Entity Relationship Diagram*

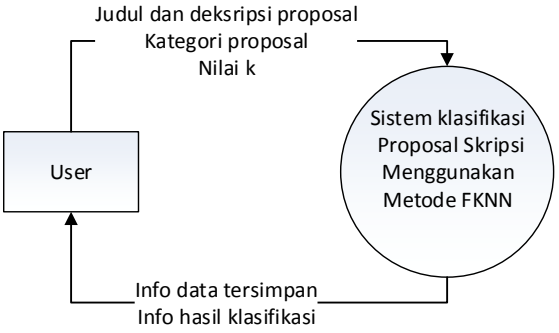
Pemodelan awal basis data untuk sistem Klasifikasi proposal skipsi menggunakan Fuzzy K-nearest Neighbour yang dibangun adalah sebagai berikut:



Gambar 4 ERD

2.2.4.2 Diagram Konteks

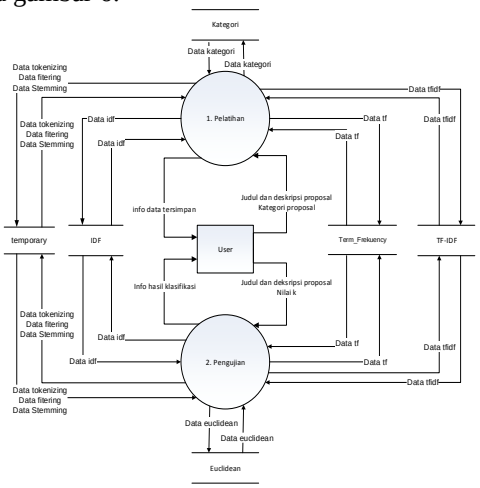
Diagram konteks menggambarkan sistem yang akan dibuat sebagai suatu entitas tunggal yang berinteraksi dengan orang maupun sistem lain.



Gambar 5 Diagram Konteks

2.2.4.3 Data Flow Diagram

DFD merupakan representasi grafik yang menggambarkan aliran informasi dan transformasi informasi yang diaplikasikan sebagai data yang mengalir dari input ke output. DFD dapat dilihat pada gambar 6.



Gambar 6 DFD

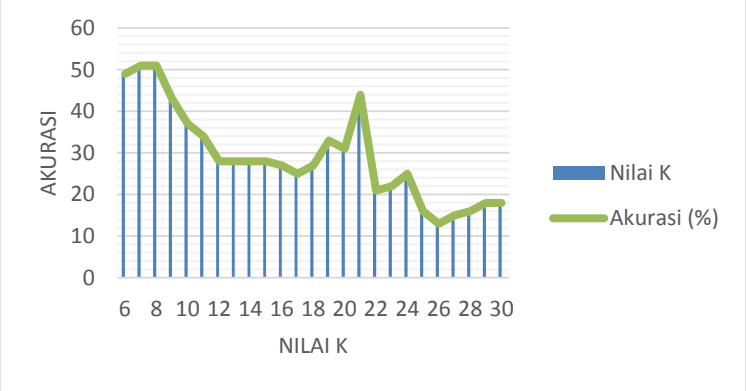
2.3 Pengujian Performansi

Pengujian performansi merupakan tahap yang memiliki tujuan untuk mengetahui performa dari metode klasifikasi yang digunakan pada sistem yang dibangun yakni metode *fuzzy* k-*nearest neighbour*. Pengujian performansi dalam penelitian ini adalah dengan mengukur tingkat keakuratan, metode pengujian yang digunakan dalam penelitian ini yakni metode confusion matrix.

Pada penelitian ini, pengujian dilakukan dengan menggunakan 1 metode. Berikut adalah pengujian yang akan dilakukan:

1. Menguji dokumen yang tidak termasuk kedalam data latih

Pengujian ini dilakukan dengan menguji data yang tidak termasuk ke dalam basis data, pengujian ini bertujuan untuk mengetahui tingkat pengenalan data uji diluar basis data terhadap data latih yang ada dalam basis data. Data latih yang digunakan sebanyak 150 buah dan data uji sebanyak 67 buah, dengan nilai K akan di isi dari nilai 6 sampai 30. Dan hasilnya dapat dilihat pada gambar 7, sebagai berikut:



Berdasarkan hasil skenario pengujian 1 yaitu pengujian data uji yang tidak terdapat dalam data latih dengan jumlah data latih sebanyak 150 buah dan nilai K yang dimulai dari 6 sampai 30, metode *fuzzy* k-*nearest neighbour* dapat mengklasifikasi dengan nilai presentase keakuratan terbesar yang diperoleh pada K=7 dan K=8 yakni sebesar 51%, untuk K=6 sebesar 49%, K=9 sebesar 43%, K=10 sebesar 37%, K=11 sebesar 34%, K=12,13,14,15 sebesar 28%, K=16,18 sebesar 27%, K=17 sebesar 25%, K=19 sebesar 33%, K=20 sebesar 31%, K=21 sebesar 44%, K=22 sebesar 21%, K=23 sebesar 22%, K=24 sebesar 25%, K=25 sebesar 16%, K=26 sebesar 13%, K=27 sebesar 15%, K=28 sebesar 16%, K=29,30 sebesar 18%. Hasil keakuratan dapat dipengaruhi oleh penggunaan nilai K yang tepat pada masing – masin kelas, jumlah data latih yang digunakan, dan kesamaan kata yang dimiliki oleh data uji dengan data latih sehingga selisih jarak perbedaan antara data uji dan data latih akan semakin kecil.

3. PENUTUP

3.1 Kesimpulan

Berasarkan hasil penelitian, analisis, pelatihan dan pengujian algoritma *Fuzzy* K-*nearest neighbour* untuk klasifikasi proposal skripsi berdasarkan kelompok keilmuan di Teknik Informatika unikom, didapatkan nilai keakuratan yang terbesar yakni dengan presentase sebesar 51% pada nilai K=7 dan K=8, maka dari itu dapat disimpulkan bahwa dalam penelitian yang dilakukan nilai K yang yang menghasilkan keakuratan terbaik yakni pada nilai K=7 dan K=8.

3.2 Saran

Dalam pembuatan tugas akhir ini, masih terdapat banyak kekurangan yang dapat diperbaiki untuk pengembangan selanjutnya. Beberapa saran yang dapat diberikan adalah sebagai berikut:

1. Menggunakan metode penghitungan jarak lain seperti *cosine similarity.*
2. Menggunakan algoritma stemming lain seperti algoritma nazief & adriani untuk stemming bahasa indonesia karena algoritma porter stemmer belum menunjukan hasil yang akurat.
3. Menggunakan aplikasi bantuan untuk membenarkan penulisan yang salah (tipo) yang terdapat pada data latih atau data uji agar membuat hasil *tokenizing*, *filtering*, dan *stemming* lebih baik.
4. Menggunakan data latih yang banyak dan beragam untuk setiap kelasnya. Sehingga presentase keakuratan klasifikasi menjadi lebih baik.
5. Deskripsi proposal skripsi digambarkan sesuai kelompok keilmuan.

4. DAFTAR PUSTAKA

[1] Farid Hamid,S.Sos.,M.Si., dan Drs.A.Rachman,MM.,M.Si., Panduan Skripsi FIKOM UMB, hal 21

[2] Prasetyo, Dimas. Muflikhah, Laili. Ridok, Achmad. "*klasifikasi genre film berdasarkan judul dan sinopsis menggunakan fuzzy k-nearest neighbour (Fuzzy K-NN)*", Prodi ilmu komputer, Program teknologi informasi dan ilmu komputer, Universitas Brawijaya.

[3] Nazir, Moh. Ph.D, "*Metode Penelitian*", Edisi Ketujuh, Bogor : Ghalia Indonesia, 2011.

[4] Pressman, Roger S, Ph.D,"*Rekayasa Perangkat Lunak*", Edisi 7,Yogyakarta: ANDI, 2012.

[5] S.A, Rosa, Shalahudin.M. "*Rekayasa Perangkat Lunak: Tertruktur dan Berorientasi Objek*". Bandung: Informatika. 2015.

[6] Sianipar, R., H. "*Panduan Praktis Pemrograman C# Bagi Pemula*". Yogyakarta: ANDI. 2014.

[7] Raymond J. Mooney. CS 391L: *Machine Learning Text Categorization.* University Of Texas at Austin, 2006.

[8] Gregorius, S. Gunawan, Ibnu. Yunono, Ferry. "*Algoritma Porter Stemmer For Bahasa Indonesia untuk pre-processing text mining berbasis Metode Market Basket Analysis*". Jurusan Teknik Informatika, UK.

[9] Lasmedi, Afuan."*Stemming Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Porter*", Prodi Teknik Informatika, Fakultas Sains dan Teknik, Universitas Jendral Sudirman.

[10] Budiharto,W., and Suhartono, D.,"*Artificial Intelligence*", Yogyakarta : Andi, 2014.

[11] Keller JM, Gray MR, Givens J.A. A Fuzzy K-Nearest Neighbor .Algorithm IEEE Transaction on Systems, Man, and Cybernetics. Vol SMC-15. No 4. 1985.

[12] Prasetyo, Eko. *Fuzzy K-Nearest Neighbor In Every Class Untuk Klasifikasi Data*. Seminar Nasional Teknik Informatika. Jurusan Teknik Informatika Fakultas Teknologi Industri Universitaas Pembangunan Nasional Veteran Jawa Timur. Pp 57-60. 2012.

[13] Wisdarianto, Ardhy. Ridok, Achmad. Rahman, Arif, Muh. "*Penerapan Metode fuzzy k-nearest neighbour (Fuzzy K-NN) untuk Pengklasifikasian Spam Email*", Prodi ilmu komputer, Program teknologi informasi dan ilmu komputer, Universitas Brawijaya.

[14] W., Teguh, Hardika. Mardji. Furqon, Tanzil, M. "*Penerapan Metode fuzzy k-nearest neighbour (Fuzzy K-NN) untuk Diagnosa Penderita Liver Berdasarkan Indian Patient Dataset(ILPD)*", Prodi ilmu komputer, Program teknologi informasi dan ilmu komputer, Universitas Brawijaya.

[15] M.,Selly, Yanita. Ridok, Achmad. Muflikhah, Laili."*Perbandingan K-Nearest Neighbor dan fuzzy k-nearest neighbour (Fuzzy K-NN) Pada Diagnosis Penyakit diabetes Melitus*", Prodi ilmu komputer, Program teknologi informasi dan ilmu komputer, Universitas Brawijaya.

[16] S., Amiratus, Rahmi. Ridok, Achmad. Muflikhah, Laili."*Penerapan Metode fuzzy k-nearest neighbour (Fuzzy K-NN) untuk menentukan Kualitas Hasil Rendemen Tanaman Tebu*", Prodi ilmu komputer, Program teknologi informasi dan ilmu komputer, Universitas Brawijaya.

[17] Puspasari, Maslikha. Dewi, Candra. Rahman, Arif, Muh. "*Prediksi Tingkat Resiko Penyakit Jantung Koroner(PJK) Menggunakan Metode fuzzy k-nearest neighbour (Fuzzy K-NN)*", Prodi ilmu komputer, Program teknologi informasi dan ilmu komputer, Universitas Brawijaya.

[18] Lestiyanto, Yudha, Dony. "*Stemming Bahasa Indonesia Sebagai Media Belajar Siswa Sekolah Menggunakan Algoritma Porter*". Sistem Informasi. Universitas Dian Nuswantoro. Semarang. 2014.

[19] Agusta, Ledy, "*Perbandingan Algoritma Stemming Porter dengan algoritma nazief & adriani untuk stemming dokumen teks bahasa indonesia"* konferensi nasional sistem dan informatika 2009. Fakultas Teknologi Informasi. Universitas Kristen Satya Wacana. KSN&109-036. 2009.

[20] Herman; Andani Achmad; dan Amil ahmad ilham, "*klasifikasi dokumen naskah dinas menggunakan algorita term frequency-inversed document frequency dan vector space model*", balai besar pengkajian dan pengembangan komunikasi dan informatika makassar, kemkominfo & Jurusan elektro, prodi informatika, fakultas teknik, UNHAS.

[21] Indriani, Aida, "*Klasifikasi Data Forum dengan menggunakan Metode Naive Bayes Classifier*", Seminar Nasional Aplikasi Teknologi Informasi(SNATI) 2014, Program studi Teknik Infomatika, STMIK PPKIA Tarakan Rahmawati. ISSN: 1907-5022.