IMPLEMENTASI ALGORITMA FUZZY K-NNC UNTUK KLASIFIKASI PROPOSAL SKRIPSI BERDASARKAN KELOMPOK KEILMUAN DI TEKNIK INFORMATIKA UNIKOM

Rozesh Eco Cevin Sagala[1], Kania Evita Dewi[2]

Teknik Informatika-Universitas Komputer Indonesia
Jalan Diaptiukur No.112-116 Bandung. 40132
Email : rozes.cevin.sagala@email.unikom.ac.id[1], kania.evita.dewi@email.unikom.ac.id[2]

ABSTRAK

Pada penelitian Yosep amin tentang klasifikasi proposal skripsi menggunakan metode *fuzzy k-nearest neighbour* (FK-NN) masih memiliki akurasi yang rendah, dimana penelitian tersebut menghasilakn akurasi sebesar 51% pada K=7 dan K=8. Pada penelitian yang dilakukan oleh Eko prasetyo mengenai "*Fuzzy K-Nearest Neighbour in every* class untuk klasifikasi data", menunjukkan bahwa metode FK-NNC lebih baik dibandingkan dengan metode FK-NN, sehingga metode FK-NNC akan diguakan untuk diimplementasikan dalam klasifikasi proposal skripsi dalam penelitian ini. Penelitian ini akan melakukan peringkasan teks terlebih dahulu sebelum dilakukan proses kalsifikasi, dimana hal ini untuk memperkecil ukuran matriks yang ada. Dalam perigkasan ini menggunakan metode *keyword extraction.* Setelah didapatkan ekstraksi teks maka hasil ekstraksi akan dimasukkan dalam data latih ataupun data uji. Parameter yang digunakan sebagai faktor penilaian yakni dilihat dari latar belakang proposal skripsi. Berdasarkan hasil pengujian dari metode FK-NNC dalam klasifikasi proposal skripsi didapatkan nilai keakuratan yang terbesar yaitu pada K=8 dan K=10 dengan nilai akurasi sebesar 77,7 %. Maka dari itu disimpulkan metode FK-NNC dapat digunakan dalam pengklasifikasian proposal skripsi.

Kata kunci : Klasifikasi, *Fuzzy K-Nearest Neighbors*, *Fuzzy* k-nnc, ekstraksi teks, peringkasan teks.

1. PENDAHULUAN

Skripsi merupakan matakuliah penutup yang harus diambil oleh mahasiswa/mahasiswi yang sedang menempuh jenjang pendidikan strata satu di perguruan tinggi. Teknik informatika UNIKOM memilki 5(lima) kelompok keilmuan yaitu kelompok A, B, C, D dan E. Setiap kelompok akan ditangani oleh dosen yang ahli di bidangnya.

Semakin berkembangnya ilmu pengetahuan dan teknologi saat ini, membuat semakin beragamnya tema-tema skripsi yang diajukan oleh mahasiswa/mahasiswi, sehingga mengakibatkan semakin besarnya irisian setiap kelompok keilmuan. Hal ini yang terkadang dapat menimbulkan kesalahan, ketika mahasiswa akan memilih kelompok keilmuan yang akan diambil dalam skripsi. Pada saat mahasiswa salah dalam memilih kelompok maka penanganan terhadap proposal skripsi seperti masukan untuk arah dalam pengerjaan skripsi tersebut akan tidak maksimal, karena proposal skripsi tidak ditangani oleh penguji yang ahli di bidangnya. Untuk membantu dalam mengidentifikasi proposal skripsi, kita dapat manfaatkan metode pengklasifikasian.

Penelitian yang berkaitan dengan pengklasifikasian dokumen sudah pernah dilakukan sebelumnya. Penelitian yang dilakukan oleh Yosep Amin mengenai "Implementasi Algoritma *Fuzzy K-nearest neighbour (Fuzzy* K-NN) untuk Klasifikasi Proposal Skripsi Berdasarkan Kelompok Keilmuan di Teknik Informatika UNIKOM", menunjukkan bahwa hasil penelitian, analisis, pelatihan dan pengujian algoritma Fuzzy K-*nearest neighbour* untuk klasifikasi Proposal didapatkan nilai keakuratan yang terbesar yakni 51% pada K=7 dan K=8 [1]. Dari hasil penelitian tersebut didapatkan akurasi yang masih rendah, dan juga masih memiliki matrik yang cukup besar sehingga dibutuhkan solusi lain agar dapat meningkatkan akurasi yang lebih optimal.

Pada penelitian ini akan menggunakan algoritma *Fuzzy K-Nearest Neighbour in every class* (FK-NNC) dalam pengklasifikasian proposal skripsi dan juga *keyword extraction* untuk memperkecil matriks data uji maupun latih [2]. Metode FK-NNC merupakan pengembangan dari Metode *Fuzzy K-Nearest Neighbour*. Penelitian yang dilakukan oleh Achmad Syaiful Aziz dalam pengklasifikasian jenis baja berdasarkan komposisi kimia dengan menggunakan metode *Fuzzy k-Nearest Neighbour in every class* menunjukkan dengan dilakukan 3(tiga*)* kali percobaan dengan menggunakan data latih yang berbeda didapatkan tingkat akurasi data rata-rata sebesar 96,7 % dan laju *error* rata-rata sebesar 3,3% [3]. Dalam penelitian Eko prasetyo menunjukkan, bahawa metode FK-NNC lebih baik dan akurat

dalam pengklasifikasian data dibandingkan metode FK-NN dan K-NN[4].

Berdasarkan penjelasan di atas, dalam penelitian ini akan dilakukan implementasi algoritma Fuzzy K-NNC dalam menyelesaikan permasalahan klasifikasi proposal skripsi. Diambil studi kasus di program studi Teknik Informatika UNIKOM.

Tujuan yang diharapkan dari penelitian ini adalah untuk mengetahui keakuratan metode *Fuzzy K-Nearest Neighbour in every class* dalam melakukan klasifikasi proposal skripsi di Teknik Informatika UNIKOM.

2. ISI PENELITIAN

2.1 Deskripsi sistem

Dalam perancangan sistem yang akan dibangun, akan dibuat sebuah aplikasi untuk mengklasifikasian proposal skripsi di Teknik Informatika UNIKOM. Dimana pada program yang akan di bangun akan menggunakan algoritma *fuzzy K-Nearest Neighbour in every clas* . Terdapat dua tahapan pada klasifikasi yang akan dilakakukan yaitu tahap data latih dan data uji. Sebelum data akan di latih maupun di uji, data masukan akan di filter/dihilangkan titik yang kurang penting, dimana hal ini dilakukan agar data *preprocessing* menghasilkan keluaran yang baik. Setelah data masukkan dilakukan proses preprocessing maka selanjutnya akan dilakukan proses kompresi sbesar 50 %, hal ini dilakukan untuk memperkecil matriks yang ada.

2.2 Peringkasan

Peringkasan teks adalah suatu kegiatan merangkum teks dari sebuah dokumen, dimana tujuan rangkuman ini untuk menyajikan gagasan utama dalam sebuah dokumen dalam ruang/penyimpanan yang lebih sedikit[5]. Pada peringkasan teks terdapat dua metode yaitu secara manual dan otomatis. Pada peringkasan teks otomatis terdapat dua prosedur yang dapat dilakukan yaitu ekstraksi dan abstraksi. Dalam proses ekstraksi menggunakan algoritma ekstraksi kata kunci dimana algoritma ini menyalin unit-unit teks yang dianggap penting atau paling informatif dati teks yang akan diringkas, sedangkan yang dilakukan pada abtraksi adalah membangun sebuah representasi senabtik internal seperti mengambil intisari dari teks, kemudian membuat ringkasan dengan menciptakan kalimat-kalimat aru yang mempresentasikan intisari dari teks. Peringkasan teks dapat dilakukan pada satu dokumen dan banyak dokumen (*multy document*).

2.3 Keyword Extraction

Pada tahap *keyword extraction* ini, dilakukan untuk untuk memperkecil matriks yang akan di simpan sebagai data latih maupun yang akan digunakan sebagai data uji. Tahap peringkasan metode *keyword extraction* terdiri dari dua tahap yaitu pertama dengan menggunakan algoritma ekstraksi kunci, lalu *selector* rasio kompresi [5]. Pada tahap pertama dokumen diberi tolak ukur dengan algoritma ekstraksi kata kunci dan kemudian menghitung TF/IDF seperti yang telah dilakukan sebelumnya, yaitu pada tahap penghitungan TF/IDF kalimat. Dari hasil Tf/Idf, dijumlah semua TF/IDF untuk setiap kalimat dan dari hasil tersebut diurutkan dari yang terbesar, setelah diurutkan maka lakukan kompresi sebesar 50% dari dokumen asli. Setelah pemilihan sebanyak 50%, proses selanjutnya adalah mengatur ulang satu kalimat daru nilai yang terbesar sampai yang paling rendah, dan teks baru ini dengan susunan kalimat baru akan menjadi teks akhir yang diringkas Kompresi dilakukan sebanyak 50% karena pada penelitian sebelumnya, didapatkan akurasi yang lebih baik dibandingkan dengan kompresi sebesar 20 dan 30%. [5].

2.4 Term Weighting/ *TF-IDF*

Tahap *Term weighting/ TF-IDF* adalah metode yang digunakan dalam menghitung bobot setiap kata. 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 mengitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) degan cara mengalikannya[6].

Term Frequency (TF) adalah menunjukkan jumlah kata yang sering muncul dalam dokumen. Jumlah kemunculan kata pada dokumen akan mempengaruhi besar dari bobot kata tersebut.

Inverse Document Frequency (IDF) adalah pengurangan dominansi *term* yang sering muncul di berbagai dokumen. terkadang kata yang banyak muncul dalam dokumen, merupakan kata yang tidak penting, tetapi kata yang muncul pada sedikit dokumen dapat dikatakan sebagai kata yang lebih penting. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu kata (*inverse document frequency*).

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

$$\boldsymbol{W_i = TF \times IDF} \qquad (2.1)$$

Dimana :

W_i = bobot kata yang ada dalam sebuah dokumen .

i = indeks kata .

TF = banyaknya kata yang muncul dalam sebuah dokumen.

IDF = *inverse document frequency* dari kata.

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

$$\boldsymbol{IDF = \log\left(\frac{D}{DF}\right)} \qquad (2.2)$$

Dimana :

IDF = hasil perhitungan dari jumlah kata keseluruhan dibagi jumlah dokumen yang terdapat kata yang di cari.
D = jumlah dokume yang ada.
DF = jumlah dokumen yang terdapat kata yang dicari.

2.5 Fuzzy K- Nearest Neighbour in Every Class

Metode *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC) adalah sebuah metode pengklasifikasian dengan menggunakan K tetangga terdekat pada setiap kelas yang ada dari sebuah data uji dan akan dijumlahkan dengan tetangga dikelas x yang ditemukan dengan menggunakan metode FK-NNC[7]. Metode FK-NNC adalah pengembangan dari metode FK-NN, dimana metode FK-NNC melakukan pengembangan dengan melakukan Cxk tetangga terdekat. Namun meskipun demikian, kerangka dari FK-NNC masih menggunakan kerangka FK-NN dimana nilai keanggotaannya berinterval [0.1] dimana Jumlah nilai keanggotaan sebuah data pada semua kelas sama dengan 1, seperti pada rumus 2.3, berikut :

$$\sum_{j=1}^{c} \boldsymbol{U_i} = \mathbf{1},\ \mathbf{0} \leq \boldsymbol{U_j} \leq \mathbf{1} \qquad (2.3)$$

Dimana
:
U_i = nilai dari keanggotaan data latih.
U_j = nilai dari data uji U_i ke kelas ke- j.

Pada tahapan selanjutnya akan dilakukan perhitungan jarak antara data *testing* dengan data *traning*, hal ini dilakukan untuk menghitung kedekatan anatar data yang akan di uji dan data latih. untuk menghitung jarak tersebut akan dilakukan dengan menggunakan metode *euclidean distance*. Berikut merupakan persamaan untuk memberikan nilai keanggotaan pada data uji W_j dengan menggunkan metode *euclidean distance*. Jarak *eucidean distance* antara vektor data uji W_j dengan vektor data latih W_i dihitung menggunakan rumus dibawah ini:

$$\|\boldsymbol{W_i} - \boldsymbol{W_j}\| = \left(\sum_{i=1}^{n} |\boldsymbol{W_i} - \boldsymbol{W_j}|^2\right)^{1/2} \qquad (2.3)$$

Dimana :

$\|W_i - W_j\|$= jarak e*uclidean* vektor data uji dan dokumen latih.
W_j = bobot hasil TF-IDF ringkasan data uji.
W_i = bobot hasil TF-IDF ringkasan data latih ke-i.
i = index *term* ke-*i*.
n = jumlah *term* keseluruhan hasil *text preprocessing*.

Setelah didapatkan jarak antara data uji dan data latih dengan metode *eucidean distance* maka selanjutnya dilakukan perhitungan akumulasi jarak uji ke setiap kelas. Dimana untuk setiap kelas akan dikalikan denan kelas K tetangga yang didapat. Selanjutnya jarak data uji U_j ke semua K tetangga dari setiap kelas akan dijumlahkan, dengan rumus :

$$\boldsymbol{S_i} = \sum_{j=1}^{K} \boldsymbol{d}(\boldsymbol{W_i}, \boldsymbol{W_j})^{-2/m-1} \qquad (2.4)$$

Dimana:
S = jarak kumulatif ke-i.
i = 1, 2, 3, 4, 5.
d = akumulasi jarak data uji W_j ke tetangga dalam kelas ke-$_j$ sebanyak C kelas.

m = pangkat bobot .
Jika pada perhitungan jarak menggunakan metode *Manhattan* maka m = 1.
Jika pada perhitungan jarak menggunakan metode *Euclidean* maka m = 2.
pada perhitungan jarak menggunakan metode *Chebyshev* m = ∞ maka.

Selanjutnya setelah mendapatkan akumulasi jarak uji ke setiap kelas akan dilakukan penjumlahan akumulasi jarak uji ke semua kelas yang ada. Dimana jarak akumulasi data uji W_j ke semua kelas disimbolkan dengan D, rumusnya sebagai berikut :

$$\boldsymbol{D} = \sum_{i=1}^{c} (\boldsymbol{S_i}) \qquad (2.5)$$

Dimana :
D = jumlah akumulasi dari semua kelas.

Tahap selanjutnya adalah melakukan perhitungan nilai keanggotaan. Agar mendapatkan nilai keanggotaan data uji W_j pada semua kelas ke-i (ada C kelas), dengan menggunakan rumus sebagai berikut :

$$\boldsymbol{U_i} = \frac{\boldsymbol{S_i}}{D} \qquad (2.6)$$

Dimana :
U_i = Keanggotaan kelas k-i.

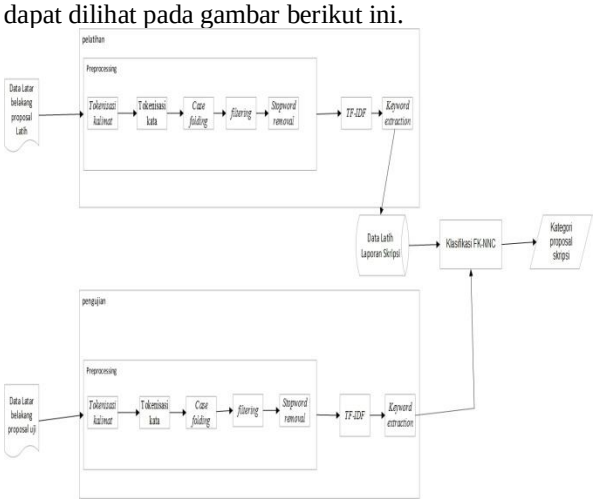
Kemudian untuk menentukan kelas hasil prediksi data uji W_j, dipilih kelas dengan nilai keanggotaan terbesar dari data U_i. Dengan rumus :

$$\boldsymbol{y'} = \mathbf{arg\,max}(\boldsymbol{U_i}) \qquad (2.7)$$

2.6 Analisis Sistem

Sistem yang akan dibangun memiliki beberapa tahapan , tahapan tersebut seperti tahap pelatihan dan tahap pengujian. Pada tahap pelatihan dan pengujian terbagi menjadi beberapa tahap lagi yaitu

tahap *preprocessing*, kemdian tahap pembobotan nilai kata, tahap ekstraksi teks dan tahap pengujian. Gambaran umum dari aplikasi yang akan dibangun dapat dilihat pada gambar berikut ini.



Gambar 1. Gambaran umum sistem

2.6 Analisis Data Masukan

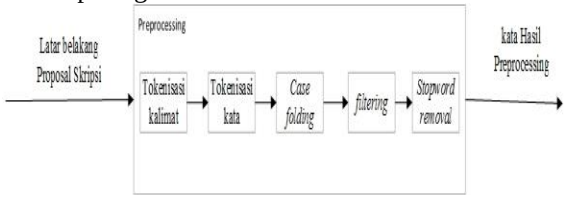
Data masukan yang di inputkan adalah dokumen latar belakang proposal skripsi yang berformat *.txt*, dimana pada setiap data masukan akan dilakukan mengecekan dan menghapusan titik, hal ini dilakukan agar data masukan dapat menghasilkan hasil preprocessing yang baik, letak titik yang di hapus seperti pada kata PT, CV, jl, no, persen dan gelar. Pada data masukan proposal skripsi ini, terdapat 5 kelas kelompok keilmuan, yaitu kelas A, B, C, D, dan E. Terdapat 55 data dokumen latar belakang propsal akan digunakan sebagai data latih dan 18 dokumen sebagai akan digunakan sebagai data uji, sehingga terdapat 73 data masukan.

2.6 Analisis Tahap Pelatihan

Tahap pelatihan dilakukan untuk mendapatkan data pembelajaran yang akan digunakan untuk data pembanding pada proses klasifikasi di tahap pengujian. Tahapan ini termasuk dalam tahap *preprocessing*, tahap *pembobotan TF-IDF* dan tahap *keyword extraction.*

2.6.1 Tahap Preprocessing

Preprocessing adalah proses pengolahan data asli sebelum diolah pada tahap selanjutnya. Masukkan data awal berupa dokumen latar belakang proposal skripsi mahasiswa teknik informatika UNIKOM. Tahapan *Preprocessing* pada penelitian ini meliputi : tokenisasi kalimat, tokenisasi kata, *case folding*, *filtering*, dan *Stopword removal* Berikut tahapan *Preprocessing* yang dilakukan dapat dilihat pada gambar 2.



Gambar 2. Tahap preprocessing

Untuk tahap *preprocessing* data akan di jelasksn pertahap sebagai berikut.

a. Tahap Tokenisasi Kalimat

Tokenisasi kaliamat digunakan untuk memisahkan teks ke dalam bentuk satuan kalimat. Tokenisasi kalimat pada sistem ini menggunkan *delimeter*(“.”). Berikut tahapan yang dilakukan dapat dilihat pada gambar 3.



Gambar 3. Tahap tokenisasi kalimat

b. Tahap tokenisasi kata

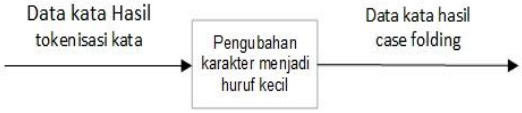
Tahap ini melakukan pengecekan dari karakter pertama sampai dengan karakter terakhir. Apabila ditemukan karakter *delimeter* seperti spasi, maka teks akan dipisahkan dan dimasukkan dalam kumpulan kalimat. Berikut tahapan yang dilakukan dapat dilihat pada gambar 4.



Gambar 4. Tahap tokenisasi kata

c. Tahap case folding

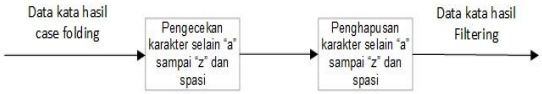
Tahap case folding merupakan proses dimana setiap kata yang ada diubah menjadi huruf kecil atau huruf kapital semua. Pada tahapan *case folding* pada penelitian ini, akan dilakukan pengubahan setiap kata menjadi huruf kecil semua.Berikut tahapan yang dilakukan, dapat dilihat pada gambar 5.



Gambar 5. Tahap case folding

d. Tahap filtering

Tahap filtering merupakan proses dimana karaker selain “a” sampai “z” dan spasi akan dihilangkan dan hanya menenrima spasi. Berikut tahapan yang dilakukan, dapat dilihat pada gambar 6.

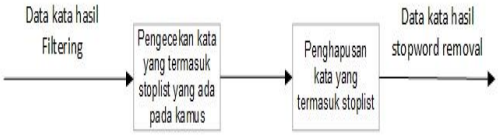


Gambar 6. Tahap filtering

e. Tahap *Stopword Removal*

Merupakan tahap menghapusan kata–kata yang kurang penting atau tidak menggambarkan isi dari dokumen yang ada.

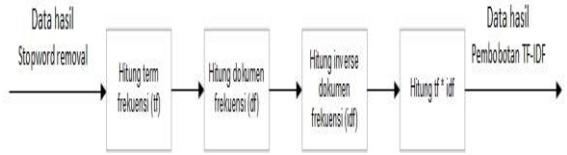
Dalam proses *Stopword Removal yang akan dilakukan, akan* menggunakan *stoplist* dari penelitian tala, namun ada penambahan dan pengurangan kata. Dimana hal ini dilakukan agar kata yang penting dalam penelitian namun terdapat dalam *stoplist* yang terdapat dalam kamus tala tidak dihapus. Berikut tahapan yang dilakukan dapat dilihat pada gambar 7.



Gambar 7. Tahap Stopword removal

2.6.2 Analisis Pembobotan kata/ TF-IDF

Tahap TF-IDF dilakukan untuk perhitungan pembobotan kata(w_i) yang ada pada dokumen. langkah pertama yaitu mencari nilai TF dan IDF dari setiap kata, untuk mendapatkan nilai IDF dilakukan perhitungan dengan persamaan (2.2). kemudian Setelah mendapatkan nilai TF dan IDF maka selajutnya menghitung pembobotan (w_i),untuk mendapatkan bobot (w_i) masing-masing dokumen pada setiap kata, Hal ini dilakukan untuk mencari bobot setiap kata dapat dilakukan dengan perhitungan persamaan (2.1). Berikut tahapan yang dilakukan dalam perhitungan pembobotan kata, dapat dilihat pada gambar 8.



Gambar 8. Pembobotan TF-IDF

2.6.3 Analisis K*eyword Extraction*

Tahap ini menggunakan peringkasan metode *keyword extraction* yang terdiri dari beberapa tahap yaitu pertama dengan menggunakan algoritma ekstraksi kunci dan *selector* rasio kompresi[10]. Pada tahap pertama, dokumen diberi tolak ukur dengan algoritma ekstraksi kata kunci dan kemudian menghitung TF/IDF seperti yang telah dilakukan sebelumnya, yaitu pada tahap penghitungan TF/IDF kalimat. Kemudian dari hasil TF/IDF, akan dijumlah semua TF/IDF untuk setiap kalimat yang ada. kemudian setelah didapatkan jumlah dari setiap kalimat (w_i), maka hasil tersebut diurutkan dari yang terbesar sampai yang terkecil. Setelah diurutkan, maka lakukan kompresi sebesar 50% dari dokumen asli. Maka dari hasi kompresi yang telah dilakukan, data hasil kompresi akan disimpan dalam database yang akan digunakan sebagai data latih.

2.7 Analisis Tahap Pengujian

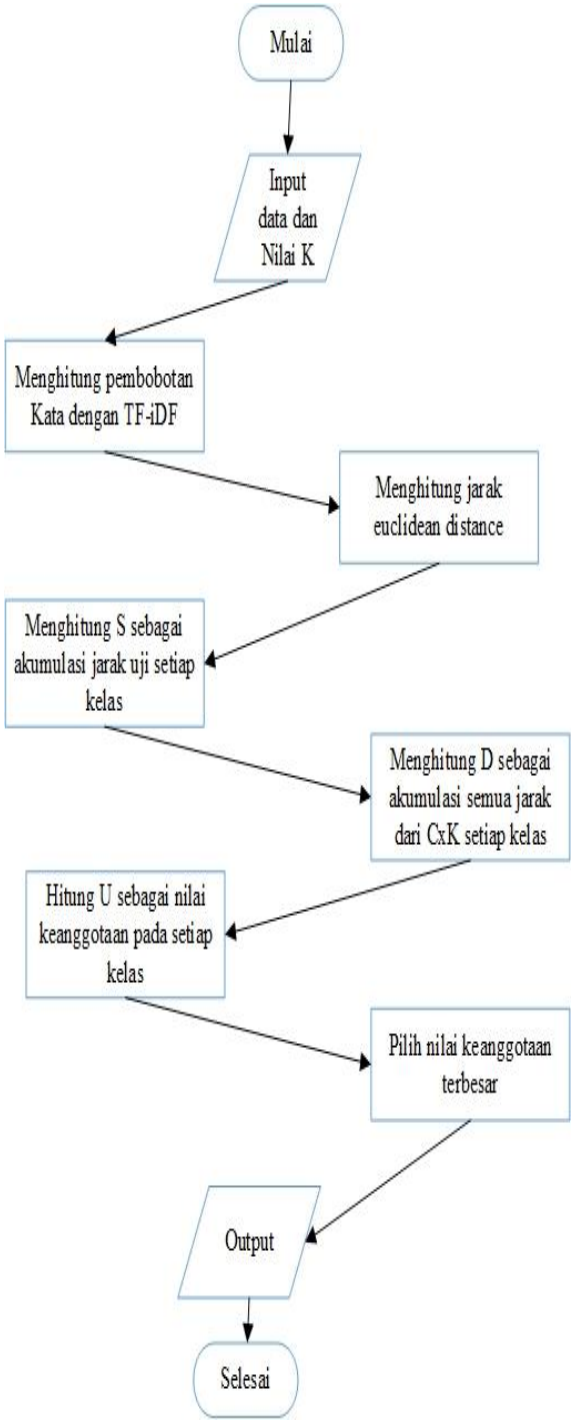
Pada tahap pengujian, langkah - langkah yang dilakukan untuk tahap pengolahan kata sama seperti yang dilakukan pada tahap pelatihan, yaitu meliputi proses *preprocessing*, pembobotan kata/ *TF-IDF* dan *keyword extraction*. Setelah didapatkan hasil *keyword extraction* yaitu berupa kompresi teks, maka hasil kompresi tersebut akan digunakan sebagai data uji. kemudian untuk mendapatkan kategori maka akan melakukan prosess klasifikasi menggunakan metode FK-NNC. Proses yang pertama dalam tahap pengujian adalah melakukan *preprocessing* dari data uji yang dimasukkan, proses *preprocessing* yang dilakukan sama seperti tahap pelatihan yaitu tahap tokenisasi kalimat, tokenisasi kata, *case folding, filtering* dan *stopword removal*. Lalu setelah dapat hasil *preprocessing* maka akan langsung dilakukan proses *keyword extraction* data uji dan melakukan pengkasifikasian.

2.7.1 Analisis Tahap Klasifikasi

Tahap klasifikasi dilakukan setelah proses pembobotan TF-IDF selesai, maka akan dilakukan klasifikasi dokumen uji dengan menggunakan metode FK-NNC. Berikut penjelasan tahapan yang ada pada algoritma FK-NNC untuk sebuah data uji, disajikan sebagai berikut :

1. Menghitung TF-IDF dari kata dokumen uji terhadap dokumen latih (2.1).
2. Menghitung jarak dokumen uji dan latih menggunakan metode *euclidean* distence (2.3)
3. Mengurutkan jarak *euclidean* distence secara ascending dengan nilai K yang sudah ditentukan.
4. Menghitung akumulasi jarak uji setiap kelas (2.4)
5. Menghitung jarak uji ke semua k tetangga dari setiap kelas (2.5)
6. Menghitung nilai keanggotaan setiap kelas (2.6)
7. Memilih hasil prediksi klasifikasi (2.7).

penjelasan diatas adalah tahapan yang dilakuka metode FK-NNC dalam melakukan klasifikasi data uji hingga mendapatkan hasil kategori. Proses diatas akan ditampilkan dalam bentuk gambar yang dapat dilihat pada gambar 9.



Gambar 9. Tahapan FK-NNC

2.8 Pengujian Akurasi

Pengujian akurasi pada penelitian klasifikasi proposal skirpsi menggunaka metode *fuzzy K-Nearest Neighbour in every class* ini dilakukan untuk mengetahui nilai akurasi untuk dapat menentukan apakah metode *fuzzy K-Nearest Neighbour in every class* cocok digunakan untuk klasifikasi proposal skripsi.

Jumlah data yang digunakan dalam pengujian ini adalah sebanyak 73 data latar belakang proposal skripsi mahasiswa Teknik Informatika UNIKOM, dimana dimana 55 data digunakan sebagai data latih dan 18 data sebagai data uji.

Dalam pengujian ini terdapat 11 data dari setiap kelompok keilmuan yang digunakan sebagai data latih, dimana terdapat 5 kelompok keilmuan. Pada pengujian ini juga akan menggunakan nilai K dari K=2 samapai dengan K=10, maka berikut adalah hasil dari pengujian latar belakang proposal skrispi yang dilakukan dengan metode *fuzzy K-Nearest Neighbour in every class.*

a. Pengujian data dengan nilai K =2.

Tabel 1. *confusionmatrix* pengujian dengan K=2

Kelompok keilmuan		PERKIRAAN				
		A	B	C	D	E
Tujuan	A	**1**	0	0	0	0
	B	0	**1**	0	0	0
	C	2	1	**3**	0	0
	D	0	1	1	**2**	1
	E	1	1	0	2	**1**

$$Akurasi = \frac{8}{18}\ x\ 100\%$$
$$Akurasi = 44{,}4\%$$

b. Pengujian data dengan nilai K =3.

Tabel 1. *confusionmatrix* pengujian dengan K=3

Kelompok keilmuan		PERKIRAAN				
		A	B	C	D	E
Tujuan	A	**1**	0	0	0	0
	B	0	**1**	0	0	0
	C	2	0	**2**	1	0
	D	0	1	2	**2**	0
	E	0	2	0	2	**1**

$$Akurasi = \frac{7}{18}\ x\ 100\%$$
$$Akurasi = 38{,}8\%$$

c. Pengujian data dengan nilai K =4.

Tabel 1. *confusionmatrix* pengujian dengan K=4

Kelompok keilmuan		PERKIRAAN				
		A	B	C	D	E
Tujuan	A	**1**	0	0	0	0
	B	0	**1**	0	0	0
	C	2	0	**3**	1	0
	D	0	0	1	**4**	0
	E	0	2	0	1	**2**

$$Akurasi = \frac{11}{18}\ x\ 100\%$$
$$Akurasi = 61{,}1\%$$

d. Pengujian data dengan nilai K =5.

Tabel 1. *confusionmatrix* pengujian dengan K=5

Kelompok keilmuan		PERKIRAAN				
		A	B	C	D	E
Tujuan	A	**1**	0	0	0	0
	B	0	**1**	0	0	0
	C	4	0	**1**	1	0
	D	0	0	1	**4**	0
	E	1	1	0	0	**2**

$$Akurasi = \frac{9}{18}\ x\ 100\%$$
$$Akurasi = 50\%$$

e. Pengujian data dengan nilai K =6.

Tabel 1. *confusionmatrix* pengujian dengan K=6

Kelompok keilmuan		PERKIRAAN				
		A	B	C	D	E
Tujuan	A	**1**	0	0	0	0
	B	0	**1**	0	0	0
	C	3	0	**1**	2	0
	D	0	0	1	**4**	0
	E	1	1	0	1	**2**

$$Akurasi = \frac{9}{18}\ x\ 100\%$$
$$Akurasi = 50\%$$

f. Pengujian data dengan nilai K =7.

Tabel 1. *confusionmatrix* pengujian dengan K=7

Kelompok keilmuan		PERKIRAAN				
		A	B	C	D	E
Tujuan	A	**1**	0	0	0	0
	B	0	**1**	0	0	0
	C	3	0	**2**	1	0
	D	0	0	0	**5**	0
	E	1	0	0	1	**3**

$$Akurasi = \frac{12}{18}\ x\ 100\%$$
$$Akurasi = 66{,}6\%$$

g. Pengujian data dengan nilai K =8.

Tabel 1. *confusionmatrix* pengujian dengan K=8

Kelompok keilmuan		PERKIRAAN				
		A	B	C	D	E
Tujuan	A	**1**	0	0	0	0
	B	0	**1**	0	0	0
	C	1	0	**4**	1	0
	D	0	0	0	**5**	0
	E	1	0	0	1	**3**

$$Akurasi = \frac{14}{18}\ x\ 100\%$$
$$Akurasi = 77{,}7\%$$

h. Pengujian data dengan nilai K =9.

Tabel 1. *confusionmatrix* pengujian dengan K=9

Kelompok keilmuan		PERKIRAAN				
		A	B	C	D	E
Tujuan	A	**1**	0	0	0	0
	B	0	**1**	0	0	0
	C	2	0	**3**	1	0
	D	0	0	0	**5**	0
	E	1	0	1	0	**3**

$$Akurasi = \frac{13}{18}\ x\ 100\%$$
$$Akurasi = 72{,}2\%$$

i. Pengujian data dengan nilai K =10.

Tabel 1. *confusionmatrix* pengujian dengan K=10

Kelompok keilmuan		PERKIRAAN				
		A	B	C	D	E
Tujuan	A	**1**	0	0	0	0
	B	0	**1**	0	0	0
	C	2	0	**3**	1	0
	D	0	0	0	**5**	0
	E	0	0	1	0	**4**

$$Akurasi = \frac{14}{18}\ x\ 100\%$$
$$Akurasi = 77{,}7\%$$

3. PENUTUP

3.1 Kesimpulan

Berdasarkana hasil penelitian yang telah dilakukan, dapat diketahui bahwa pengklasifikasian proposal skripsi dengan menggunakan metode *Fuzzy K-Nearest Neighbour in every class (*FK-NNC) memiliki akurasi yang lebih baik dibandingkan dengan menggunakan metode *Fuzzy K-Nearest Neighbour*. Pengujian pengklasifikasian proposal skripsi dengan menggunakan metode FK-NNC dengan 73 data, dimana 55 data digunakan sebagai data latih dan 18 data sebagai data uji, menghasilkan akurasi terbesar 77,7% pada K=8 dan K=10.

Hasil keakuratan dapat dipengaruhi oleh penggunaan nilai K yang tepat pada masing– masing kelas dan kesamaan kata yang dimiliki oleh data uji dengan data latih.

3.2 Saran

Dari keterbatasan yang ada, saran untuk pegembangan algoritma *Fuzzy K-Nearest Neighbour in every class (*FK-NNC) pada klasifikasi proposal skripsi adalah untuk penelitian selanjutnya dapat menggunakan aplikasi bantuan untuk melakukan pengecekan otomatis pada penulisan(tipo) ataupun dalam penghapusan titik yang tidak perlu, agar pada data latih atau data uji menghasilkan proses *preprocessing* yang lebih baik

DAFTAR PUSTAKA

[1] Amin, Yosep. 2016. "Implementasi Algoritma *Fuzzy K-Nearest Neighbour (Fuzzy K-NN)* Untuk Klasifikasi Proposal Skripsi Berdasarkan Kelompok Keilmuan Di Teknik Informatika UNIKOM*".Skripsi*. Universitas Komputer Indonesia.

[2] Darmawan, Rizki; Wahono, Romi Satria;, "Jornal of Intelligent System," *Hybrid Keyword Extraction Algorithm and Cosine Similarity for Improving Sentences Cohesion in Text Summarization*, vol. 1, Desember 2015.

[3] Rosyid, Harunur; Prasetyo, Eko; Agustin, Soffiana;, "seminar nasional teknologi informai dan multi media," *PERBAIKAN AKURASI FUZZY K-NEAREST NEIGHBOR IN*, Januari 2013.

[4] Achmad Syaiful Aziz. 2017. "*APPLICATION OF CLASSIFICATION OF STEEL BY USING CHEMICAL COMPOSITION FK-NNC (FUZZY K-NEAREST NEIGHBORIN EVERY CLASS)*".*Skripsi.* Universitas Muhammadiyah Gresik.

[5] Darmawan, Rizki; Wahono, Romi Satria;, "Jornal of Intelligent System," *Hybrid Keyword Extraction Algorithm and Cosine Similarity for Improving Sentences Cohesion in Text Summarization*, vol. 1, Desember 2015.

[6] Suliantoro, Deddy Wijaya;, *INTEGRASI PEMBOBOTAN TF IDF PADA METODE K-MEANS UNTUK*, April 2012.

[7] Kembang Rinci Hapsari and Yunus Juli Santoso, "Prosiding Seminar Nasional Manajemen Teknologi XXII," *STEMMING ARTIKEL BERBAHASA INDONESIA DENGAN PENDEKATAN CONFIX-STRIPPING*, Januari 2015.