

USULAN TUGAS AKHIR

1. IDENTITAS PENGUSUL

NAMA : Anita Susanti
NRP : 5110100145
DOSEN WALI : Ir. FX. Arunanto, M.Sc
DOSEN PEMBIMBING : 1. Umi Laili Yuhana, S.Kom, M.Sc
2. Nurul Fajrin Ariyani, S.Kom, M.Sc

2. JUDUL TUGAS AKHIR

“Rancang Bangun Modul Pengelompokan Dokumen Pada Sistem Manajemen Pengetahuan Kepegawaian”

3. LATAR BELAKANG

Pada saat ini sudah banyak sekali dokumen-dokumen yang dipublikasikan di internet atau menggunakan berbagai media lainnya. Orang atau institusi seperti perpustakaan yang membutuhkan dapat mencari dokumen tersebut dari mana saja. Setelah mencari dokumen tersebut, kemudian perlu dipilah-pilah agar penyimpanan dokumen lebih teroganisir.

Dokumen kepegawaian sangat erat hubungannya dengan pegawai institusi, baik itu struktural maupun fungsional. Pegawai struktural meliputi pegawai tata usaha, pustakawan, pegawai badan akademik dan lain sebagainya. Sedangkan yang dimaksud pegawai fungsional yaitu dosen atau tenaga pendidik. Dokumen informasi yang dibutuhkan antara pegawai struktural dan fungsional pastilah berbeda. Apalagi untuk masalah kenaikan pangkat, pasti setiap pegawai membutuhkan informasi dengan dokumen-dokumen terkait. Akan tetapi, pegawai kesulitan untuk mendapatkan

dokumen terkait. Disebabkan oleh beberapa hal antara lain: kurangnya informasi, perubahan peraturan yang mendadak.

Untuk mengatasi masalah tersebut, maka akan digunakan metode klasifikasi untuk melakukan pengelompokan dokumen kepegawaian tersebut ke dalam sejumlah kategori. Klasifikasi dalam data *mining* didefinisikan sebagai metode pembelajaran data untuk memprediksi nilai dari sekelompok atribut. Algoritma klasifikasi akan menghasilkan sekumpulan aturan yang disebut *rule* yang akan digunakan sebagai indikator untuk dapat memprediksi kelas dari data yang ingin diprediksi. Tujuan dari algoritma klasifikasi adalah untuk menemukan relasi antara beberapa variabel yang tergolong dalam kelas yang sama. Relasi tersebut akan digambarkan dengan aturan-aturan agar dapat memprediksi kelas dari data yang atributnya sudah diketahui [1].

Tujuan dari tugas akhir ini adalah untuk implementasi sistem klasifikasi dokumen kepegawaian dengan banyak kategori sehingga dapat mempermudah dalam pencarian dokumen kepegawaian dan mengetahui tingkat akurasi hasil klasifikasi dengan metode TF-IDF dan Fuzzy K-Means dalam mengklasifikasikan dokumen kepegawaian.

4. RUMUSAN MASALAH

Beberapa rumusan masalah pada proposal tugas akhir ini yaitu:

1. Bagaimana membuat aplikasi yang dapat melakukan ekstraksi kata kunci secara otomatis dari dokumen?
2. Bagaimana membuat aplikasi yang dapat melakukan pengkategorian dokumen secara otomatis kedalam kategori yang sudah ditentukan sebelumnya?

5. BATASAN MASALAH

Permasalahan yang dibahas dalam tugas akhir ini memiliki beberapa batasan, yaitu:

1. Data masukan berupa *file* PDF.
2. Data yang digunakan untuk pembelajaran yaitu dokumen kepegawaian dan data yang didapatkan dari website yang memuat dokumen kepegawaian.
3. Sistem informasi dibangun dengan menggunakan bahasa pemrograman PHP dengan Framework Code Igniter dan *database* MySQL.
4. Kategori dokumen sudah ditentukan antara lain: Jabatan, Golongan, Dosen dan Karyawan.

6. TUJUAN PEMBUATAN TUGAS AKHIR

Beberapa tujuan dari tugas akhir ini yaitu:

1. Membuat aplikasi yang mampu melakukan ekstraksi kata kunci secara otomatis dari dokumen.
2. Membuat aplikasi pengkategorian dokumen kepegawaian dengan beberapa kategori.

7. MANFAAT TUGAS AKHIR

Dengan dibangunnya sistem pengkategorian dokumen ini diharapkan dapat mempermudah pegawai melakukan pencarian informasi dari dokumen kepegawaian yang telah dikategorikan.

8. TINJAUAN PUSTAKA

8.1. Pembobotan TF-IDF

Pembobotan TF-IDF adalah metode yang digunakan dalam membuat definisi sebuah dokumen dalam bentuk model ruang vektor. Perhitungan TF-IDF menggabungkan dua buah perhitungan yaitu perhitungan *term frequency* (TF) dan perhitungan *inverse document frequency* (IDF) [2].

Cara perhitungan TF-IDF adalah sebagai berikut:

Misalkan n adalah jumlah dokumen yang terkumpul dan kata kunci k_i muncul pada sejumlah n_i dokumen. Kemudian misalkan $f_{i,j}$ adalah banyaknya kata kunci k_i yang ditemukan pada dokumen d_j . Maka $TF_{i,j}$, yaitu *term frequency* dari kata kunci k_i pada dokumen d_j dapat didefinisikan pada Persamaan 1.

$$TF = \frac{J_{i,j}}{\max_z f_{z,j}} \quad (1)$$

Karena tidak semua kata kunci yang muncul pada tiap dokumen bisa digunakan untuk memisahkan antara kata-kata yang benar-benar relevan dengan yang tidak, maka digunakanlah perhitungan *Inverse Document Frequency* (IDF) yang nantinya digabungkan dengan perhitungan TF. Nilai IDF dari sebuah kata kunci k_i dapat didefinisikan pada Persamaan 2.

$$IDF = \log \frac{n}{n_i} \quad (2)$$

Kemudian, bobot TF-IDF dari sebuah kata kunci k_i pada dokumen d_j dapat didefinisikan pada Persamaan 3.

$$W_{ij} = TF_{ij} \times IDF_i \quad (3)$$

Setelah pembobotan dilakukan pada tiap kata kunci, konten dari sebuah dokumen yang direpresentasikan dalam bentuk model ruang vektor dapat didefinisikan pada Persamaan 4.

$$\text{content}(d_i) = \langle W_{1j}, W_{2j}, \dots, W_{kj} \rangle \quad (4)$$

Pembobotan TF-IDF merefleksikan seberapa penting sebuah kata pada dokumen dalam kumpulan dokumen. TF-IDF sering digunakan sebagai faktor bobot dalam pencarian informasi dan *text mining*. Variasi dari pembobotan TF-IDF sering digunakan oleh mesin pencarian sebagai media utama dalam menghitung dan memberi peringkat terhadap tingkat relevansi dari dokumen berdasarkan *query* pengguna. TF-IDF dapat digunakan untuk menyaring *stop-words* dalam berbagai bidang subjek termasuk peringkasan teks dan klasifikasi.

8.2 Klasifikasi Fuzzy K-Means

Klasifikasi Fuzzy adalah metode klasifikasi data yang mengizinkan atribut-atribut data untuk menerapkan nilai keanggotaan pada objek sehingga sebuah objek dapat dianggap sebagai anggota parsial dari sebuah kelas. Keanggotaan kelas biasanya didefinisikan dengan skala kontinyu dari nol hingga satu dimana nol berarti objek bukan merupakan anggota dari kelas dan satu berarti objek memiliki keanggotaan penuh pada kelas tersebut [3]. Dengan menggunakan klasifikasi Fuzzy, sebuah data dapat memiliki lebih dari satu kelas dengan nilai keanggotaan yang berbeda-beda.

Klasifikasi termasuk ke dalam area pengenalan pola dan mesin pembelajaran. Sifat dari klasifikasi Fuzzy adalah sebagai berikut:

1. *Soft labelling*

Asumsi standar dari pengenalan pola adalah bahwa kelas-kelas yang ada bersifat *mutually exclusive* atau saling lepas. Hal ini tidak berlaku pada klasifikasi Fuzzy.

2. *Interpretability*

Klasifikasi Fuzzy seringkali didesain transparan yang berarti bahwa tahapan dan pernyataan logika yang mengarah ke proses memprediksi kelas dapat dilacak dan dipahami.

3. Data terbatas dan memerlukan keahlian khusus

Tidak semua data dapat diolah dengan menggunakan klasifikasi Fuzzy. Contoh data yang dapat diolah dengan klasifikasi Fuzzy antara lain data penyakit langka, aktivitas teroris, dan bencana alam. Klasifikasi Fuzzy dapat dibangun dengan menggunakan opini dari pakar, data, atau keduanya.

Dalam pengerjaan tugas akhir ini, metode klasifikasi Fuzzy yang digunakan menerapkan konsep K-Means. Dalam statistik dan mesin pembelajaran, pengelompokan K-Means merupakan metode analisis kelompok yang mengarah pada pemartisian N objek pengamatan ke dalam K kelompok (*cluster*)

dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan mean (rata-rata) terdekat, mirip dengan algoritma *Expectation-Maximization* untuk *Gaussian Mixture* di mana keduanya mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan yang dilakukan oleh kedua algoritma [4]. Aplikasi yang dibangun akan menggunakan metode klasifikasi untuk membangun model yang nantinya akan digunakan untuk menentukan kelas dari data. Sedangkan konsep K-Means akan digunakan untuk mengukur tingkat kedekatan sebuah data dengan kelompok tertentu pada model. Nilai tingkat kedekatan sebuah data dengan kelompok tertentu pada model tersebut akan digunakan sebagai derajat keanggotaan data dalam kelompok tertentu sehingga sebuah data dapat terkelompokkan ke dalam lebih dari satu kategori tertentu dilihat dari derajat keanggotaan data tersebut.

8.3 Text Mining

Text mining merupakan jenis pengolahan aktivitas pengolahan teks yang bertujuan untuk mencari, mengorganisir, dan menganalisa informasi. Sedangkan pengenalan formal text mining adalah penciptaan informasi baru yang belum dimengerti dalam suatu koleksi dokumen. Informasi baru tersebut didefinisikan sebagai tren, pola, atau hubungan yang tidak dapat dengan mudah dikumpulkan hanya dengan membaca text [5].

Stemming adalah salah satu contoh pengolahan *text mining*. *Stemming* digunakan dalam penguraian kata menjadi kata dasarnya. Pada *stemming* Bahasa Indonesia, proses *stemming* berjalan dengan membandingkan kata dengan data-data kata dasar yang telah kita kenali, jika belum terdapat kata yang cocok maka dilanjutkan dengan mengidentifikasi imbuhan yang terdapat dalam suatu kata. Imbuhan-imbuhan dapat berupa awalan maupun akhiran. Setelah imbuhan tersebut dikenali. Proses dilanjutkan dengan penghapusan imbuhan tersebut sehingga kata yang ditinjau berubah menjadi kata dasarnya [6].

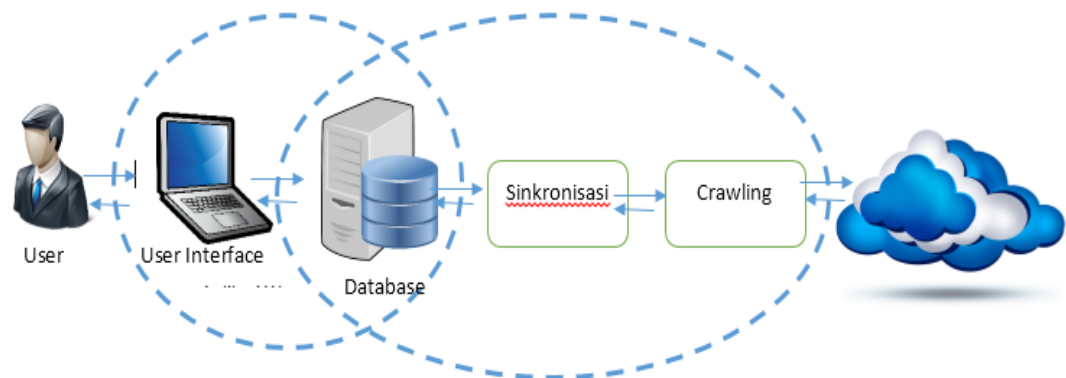
8.4 CodeIgniter

CodeIgniter adalah salah satu *framework* yang dapat digunakan untuk mengembangkan *web* dengan menggunakan bahasa pemrograman PHP. CodeIgniter menggunakan *pattern* MVC dalam pengembangan web.

9. RINGKASAN ISI TUGAS AKHIR

Sistem yang akan dibangun berbasis *web* dengan bahasa pemrograman PHP dan menggunakan *framework* CodeIgniter. Proses pengerjaan Tugas Akhir ini memuat tiga proses utama yaitu:

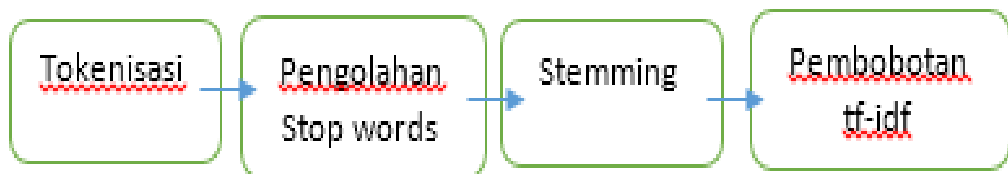
1. *Crawling* Dokumen



Gambar 1. Proses *Web Crawler*

Pada Gambar 1 dijelaskan alur dari proses *crawling* web. Pada sistem ini akan mengambil data dengan melakukan proses *crawling* ke URL yang berisi dokumen kepegawaian. Lalu dokumen yang didapat dari proses *crawling* akan melalui proses pengolahan data dan klasifikasi kemudian disimpan ke dalam *database server*.

2. Pengolahan Dokumen



Gambar 2. Proses Pengolahan Dokumen

Pada tahap pengolahan dokumen yang masuk terdapat beberapa tahapan yaitu:

a. Tokenisasi

Tokenisasi adalah proses memisahkan kata-kata dalam kalimat berdasarkan *delimiter* seperti spasi, tanda hubung, garis miring, dan lain-lain.

b. Penghilangan *stop-words*

Penghilangan *stop-words* adalah proses membuang kata-kata yang tidak dapat mewakili dokumen seperti kata 'yang', 'dan', 'dari', 'namun', dan lain-lain.

c. Stemming

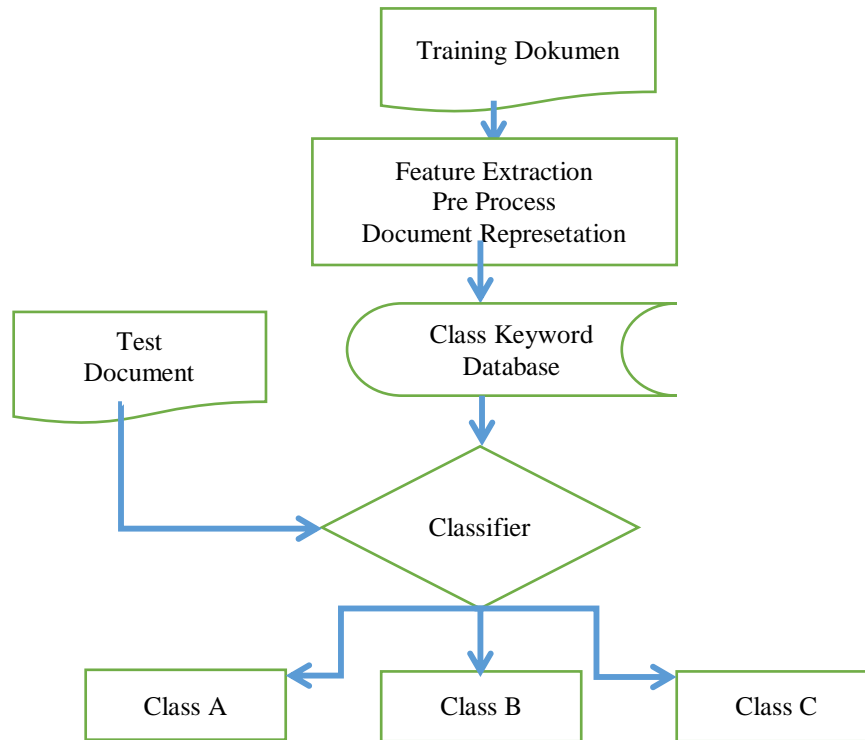
Stemming adalah proses membentuk kata dasar dari kata berimbuhan. Proses ini dilakukan dengan membuang imbuhan, sisipan, dan akhiran pada kata. Contohnya kata menyapu diubah menjadi nyapu (membuang imbuhan me-).

d. Pembobotan TF-IDF

Pembobotan TF-IDF digunakan untuk menghitung tingkat kepentingan dari sebuah kata dalam dokumen. Perhitungan ini digunakan untuk menentukan kata kunci dari dokumen.

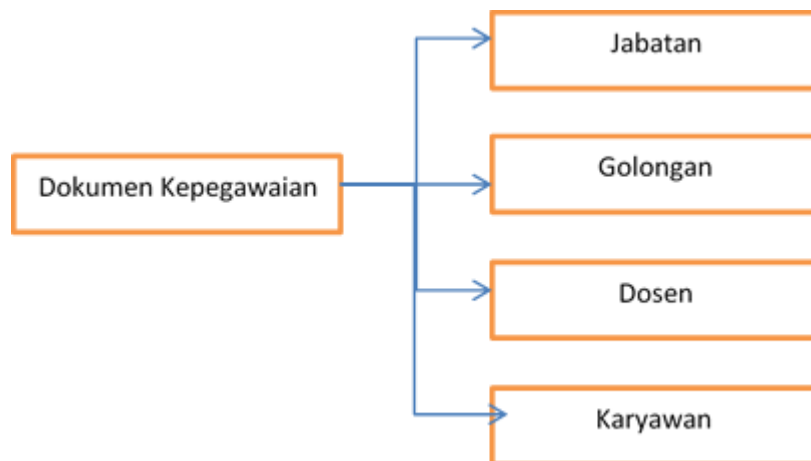
3. Klasifikasi Dokumen

Dalam hal ini metode yang akan digunakan untuk klasifikasi dokumen adalah *fuzzy k-means*. Adapun proses dari metode *fuzzy k-means* pada sistem dijelaskan pada Gambar 3.



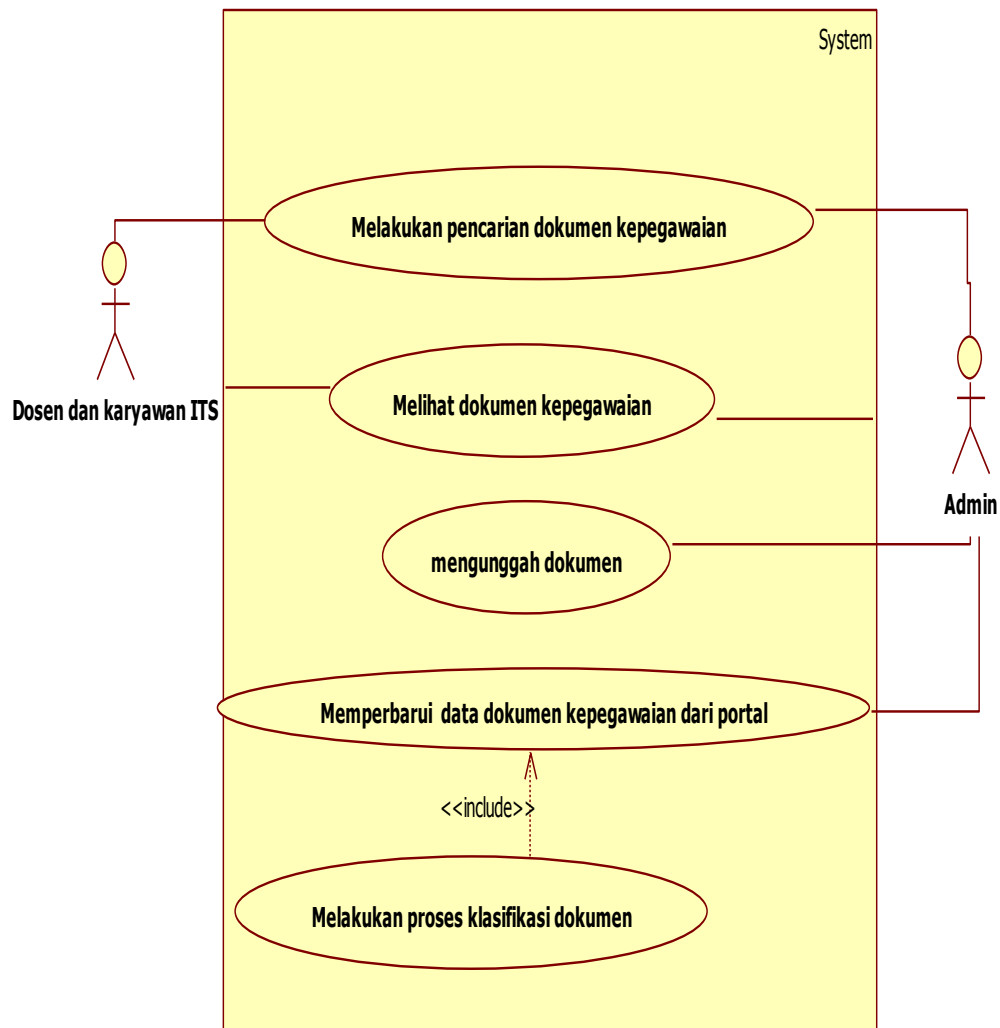
Gambar 3. Proses Klasifikasi Dokumen

Klasifikasi Fuzzy digunakan untuk mengelompokkan dokumen yang memiliki kesamaan topik. Beberapa kelas klasifikasi yang akan digunakan dijelaskan pada Gambar 4.



Gambar 4. Kategori Dokumen

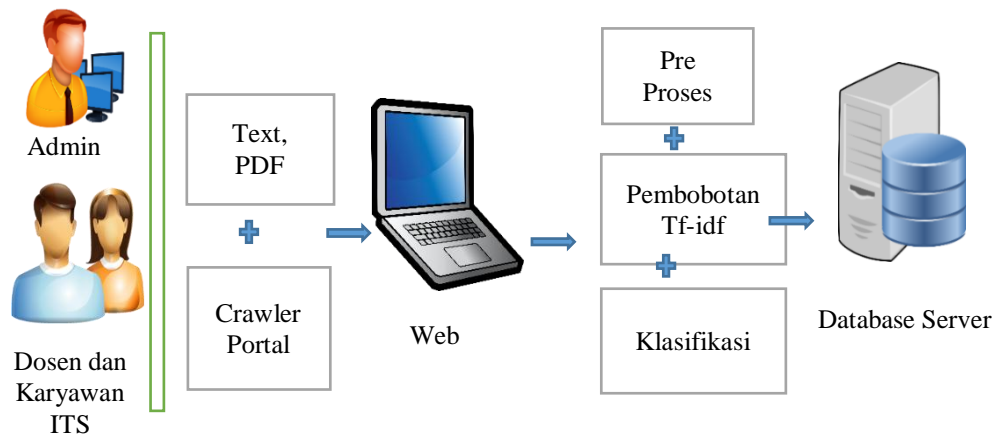
Aplikasi ini dibuat berdasarkan kebutuhan pengguna. Dari kebutuhan pengguna tersebut dapat digambarkan pada diagram *use case* dalam Gambar 5.



Gambar 5. Diagram *Use Case*

Dari diagram tersebut, fitur yang dapat diakses oleh pengguna yaitu:

- a. Pencarian dokumen kepegawaian
Fitur ini akan membantu pengguna untuk mencari dokumen kepegawaian yang dibutuhkan.
- b. Menampilkan dokumen kepegawaian kedalam beberapa kategori
Fitur ini memudahkan pengguna untuk melihat dokumen yang dibutuhkan sesuai dengan kategori yang ada yaitu: pangkat, jabatan, dosen dan karyawan.
- c. Mengunggah dokumen
Fitur ini memudahkan admin untuk mengunggah dokumen kepegawaian yang baru untuk menambah dokumen ke dalam sistem.



Gambar 6. Arsitektur Sistem

Gambar 6 menunjukkan arsitektur sistem yang akan dibangun. Pengguna memasukkan kata kunci pencarian pada kotak teks yang disediakan dimana kata kunci pencarian yang dimaksud adalah kriteria pencarian yang didasarkan pada pengetahuan dokumen kepegawaian. Pengguna (admin) dapat mengunggah dokumen kepegawaian yang baru ke dalam sistem. Serta sistem akan melakukan *crawling* dokumen kepegawaian dari portal. Semua masukan dari luar sistem akan diproses dengan beberapa tahapan yaitu: pra proses, pembobotan TF-IDF dan klasifikasi dokumen. Hasil dari pemrosesan akan disimpan kedalam *database server*.

10.METODOLOGI

Metode yang akan dilakukan dalam tugas akhir ini memiliki beberapa tahapan, yaitu:

a. Penyusunan proposal tugas akhir

Proposal tugas akhir ini berisi tentang deskripsi pendahuluan dari tugas akhir yang akan dibuat. Pendahuluan ini terdiri atas hal yang menjadi latar belakang diajukannya usulan tugas akhir, rumusan masalah yang diangkat, batasan masalah untuk tugas akhir, tujuan dari pembuatan tugas akhir, dan manfaat dari hasil pembuatan tugas akhir. Selain itu dijabarkan pula tinjauan pustaka yang digunakan sebagai referensi pendukung pembuatan tugas akhir. Sub bab metodologi berisi penjelasan mengenai tahapan penyusunan tugas akhir mulai dari penyusunan proposal hingga penyusunan buku tugas akhir. Terdapat pula sub bab jadwal kegiatan yang menjelaskan jadwal pengerjaan tugas akhir.

b. Studi literatur

Pada studi literatur ini, akan dipelajari sejumlah referensi yang diperlukan dalam pembuatan aplikasi yaitu mengenai dokumen kepegawaian, dokumen kepegawaian dari *website*, *text mining*, pembobotan TF-IDF, dan klasifikasi Fuzzy.

c. Analisis dan desain perangkat lunak

Aktor dari aplikasi ini adalah dosen dan karyawan ITS. Fitur yang terdapat pada aplikasi ini adalah:

1. Mengunggah dokumen
2. Menampilkan dokumen
3. Mencari dokumen
4. Mengkategorikan dokumen

d. Implementasi perangkat lunak

Aplikasi ini akan dibangun dengan bahasa pemrograman PHP dengan *framework* CodeIgniter. Aplikasi akan dibangun dengan menggunakan *Integrated Development Environment* (IDE) Netbeans IDE 7.3.1 dan MySQL sebagai *Relational Database Management System* (RDBMS).

e. Pengujian dan evaluasi

Pengujian dari aplikasi ini akan dilakukan pengujian tingkat akurasi, presisi, dan *recall* digunakan untuk mengukur kemampuan aplikasi dalam melakukan pengelompokan dokumen, seberapa akurat dan presisi dokumen dikelompokkan ke dalam kategori yang sesuai yang dibuat dalam membantu dosen dan karyawan ITS.

f. Penyusunan Buku Tugas Akhir

Pada tahap ini dilakukan penyusunan laporan yang menjelaskan dasar teori dan metode yang digunakan dalam tugas akhir ini serta hasil dari implementasi aplikasi perangkat lunak yang telah dibuat. Sistematika penulisan buku tugas akhir secara garis besar antara lain:

1. Pendahuluan
 - a. Latar Belakang
 - b. Rumusan Masalah
 - c. Batasan Tugas Akhir
 - d. Tujuan
 - e. Metodologi
 - f. Sistematika Penulisan
2. Tinjauan Pustaka
3. Desain dan Implementasi
4. Pengujian dan Evaluasi
5. Kesimpulan dan Saran
6. Daftar Pustaka

11. JADWAL KEGIATAN

Jadwal pengerjaan tugas akhir ditunjukkan pada Tabel 1.

Tabel 1. Jadwal Pengerjaan Tugas Akhir

Tahapan	2014																		
	Februari					Maret				April				Mei				Juni	
Penyusunan Proposal																			
Studi Literatur																			
Implementasi																			
Pengujian dan evaluasi																			
Penyusunan buku																			

12. DAFTAR PUSTAKA

- [1] J. A. Sofyan, "Public Media Learning," November 2012. [Online]. Available: <http://www.publicmedialearning.com/discussion/view/241/klasifikasi-dalamdatamining>. [Accessed 12 Februari 2014].
- [2] G. Ramadhan, in *Rekomendasi Konten Berdasarkan Profil Pribadi Pengguna Yang Dinamis*, Surabaya, ITS Surabaya, 2013.
- [3] D. GIS, "Dictionary GIS," [Online]. Available: <http://support.esri.com/en/knowledgebase/GISDictionary/term/fuzzy%20classification>. [Accessed 10 Februari 2014].
- [4] E. Prasetyo, *Data Mining, Konsep dan Aplikasi Menggunakan Matlab*, Indonesia: C. V. Andi Offset, 2012.
- [5] R. a. S. J. Feldman, *The Text Mining Handbook*, New York: Cambridge University Press, 2006.
- [6] F. Z. Tala, *A Study of Stemming Effect on Information retrieval in Bahasa Indonesia*, Institute for Logic, Language and Computation, Universiteit van Amsterdam The Netherlands, 2003.