BAB III

ANALISIS DAN PERANCANGAN

III.1 Analisis Sistem

Analisis sistem adalah penelitian atas sistem yang telah ada de ngan tujuan untuk perancang system yang baru atau diperbarui. Tahap analisis sistem ini merupakan tahap yang sangat kritis dan sangat penting, karena kesalahan di dalam tahap ini akan menyebabkan juga kesalahan di tahap selanjutnya. Tugas utama analis sistem dalam tahap ini adalah menemukan kelemahan-kelemahan dari sistem yang berjalan sehingga dapat diusulkan perbaikannya. Dalam membangun perangkat lunak ini dilakukan beberapa tahap analisis yaitu :

1) Analisis Masalah
2) Analisis Sistem yang Sedang Berjalan
3) Analisis Text *Mining*
4) Analisis Kebutuhan Non-Fungsional
5) Analisis Kebutuhan Fungsional

III.1.1. Analisis Masalah

Analisis masalah merupakan tahapan awal yang dilakukan dalam tahap analisis sistem. Masalah dapat didefinisikan sebagai sesuatu hal yang yang ingin dipecahkan, masalah inilah yang menyebabkan tujuan suatu sistem tidak dapat dicapai. Oleh karena itu perlu didefinisikan terlebih dahulu masalah yang terjadi pada sistem saat ini.

Berdasarkan hasil wawancara dengan pihak Pusat Penelitian Informatika – LIPI, didapat beberapa permasalahan sebagai berikut :

1) Pengelompokan dokumen artikel ilmiah yang dilakukan pihak Pusat Penelitian Informatika – LIPI masih digabung dalam satu kelompok penyimpanan digital.
2) Pihak Pusat Penelitian Informatika – LIPI kesulitan dalam mengakses informasi, mengkaji artikel ilmiah, memetakan, memonitoring, mengarahkan dan mengevaluasi informasi yang terkandung di dalam tumpukan dokumen artikel ilmiah tersebut disebabkan karena belum adanya sistem untuk memfilter artikel ilmiah berdasarkan karakteristik masing – masing dokumen.

III.1.2. Analisis Sistem yang Sedang Berjalan

Analisis sistem yang sedang berjalan adalah salah satu proses yang ada pada analisis sistem yang bertujuan untuk memberikan gambaran tentang bagaimana cara kerja dari sistem yang berjalan.

Proses pengelompokan yang dilakukan pihak Pusat Penelitian Informatika – LIPI dilihat pada prosedur sebagai berikut:

1. Pusat Penelitian Informatika – LIPI menerima artikel – artikel ilmiah dari kalangan masyarakat dan tim riset yang ada di Pusat Penelitian Informatika – LIPI.
2. Pusat Penelitian Informatika – LIPI mempublikasikan artikel ilmiah melalui versi cetak.
3. Untuk arsip Pusat Penelitian Informatika – LIPI, artikel ilmiah disimpan dalam satu penyimpanan digital dalam format *.Pdf.
4. Artikel ilmiah terbit dua kali dalam setahun yaitu dibulan Mei dan November
5. Artikel ilmiah mengkaji masalah yang berhubungan dengan Informatika, Sistem Kendali dan Komputer

III.1.3. Analisis Karakteristik Sumber Data

Analisis karakteristik sumber data merupakan kegiatan yang dilakukan dalam menganalisis seluruh data yang akan digunakan. Dokumen yang akan digunakan sebagai data inputan dalam perangkat lunak pada pengelompokan artikel ilmiah di Pusat Penelitian Informatika – LIPI adalah arsip artikel ilmiah yang berformat *.pdf

Datasets dalam tahapan analisis ini berjumlah lima dokumen berformat *.pdf sebagai contoh perhitungan yang diambil dari jurnal ilmiah di Pusat Penelitian Informatika – LIPI. Dokumen artikel ilmiah yang berformat *.pdf akan ditransformasikan kedalam format *.txt dan selanjutnya subjudul abstrak yang ada di setiap dokumen artikel ilmiah berformat *.txt akan di ekstrak untuk memudahkan dalam tahapan selanjutnya. Mengingat bahwa abstrak adalah rangkuman yang terkandung disetiap artikel ilmiah tersebut.

Tahapan dalam pengambilan abstrak pada artikel ilmiah adalah sebagai berikut:

1. Dokumen dalam format *.Pdf ditransformasikan ke format *.Txt.
2. Dalam format *.Pdf, artikel ilmiah mempunyai bagian judul artikel ilmiah dan beberapa subjudul seperti abstrak.
3. Pada format *.Txt bagian subjudul abstrak artikel ilmiah akan diekstrak.
4. Semua kata yang ada pada isi subjudul abstrak akan di ekstrak sampai ketemu subjudul baru yaitu Kata Kunci.

Pada tahapan ini akan diimplementasikan secara umum dengan menggunakan pemrograman JAVA. Sistem akan dibangun menggunakan PdfBox sebagai java library. PdfBox memiliki fungsi ekstraksi dan transformasi sebagai dokumen inputan untuk text mining.

III.1.4. Analisis Text Mining

Text Clustering adalah proses unsupervised learning (proses pembelajaran sendiri) yang mengelompokkan kumpulan dokumen berdasarkan hubungan kemiripannya dan memisahkannya ke dalam beberapa kelompok**[5].**

III.1.4.1. Analisis Preprocessing

Tahapan *preprocessing* adalah tahapan awal sebelum dilakukan proses *clustering*, tahapan ini dilakukan untuk mengubah suatu dokumen kedalam format yang sesuai agar dapat diproses dalam bentuk yang tepat dan dapat diproses pada tahapan selanjutnya. Penelitian ini menggunakan tiga tahap untuk *preprocessing*, yaitu: *tokenization*, *stopword*, dan *stemming*.

Berikut contoh setiap tahapan yang diambil dari abstrak setiap dokumen.

1. ***Tokenization***, merupakan tahapan penguraian string teks menjadi *term* atau kata.

Abstrak Dokumen1: "Pembajakan hak cipta terhadap konten musik digital masih menjadi masalah besar dalam industry musik. Hal tersebut dikarenakan mudahnya proses pembajakan dan kemudahan distribusi konten digital melalui internet. Isu perlindungan hak cipta menjadi hal sangat penting untuk diterapkan dalam industri musik. Digital right management (DRM) dan audio watermarking adalah cara yang bisa diterapkan untuk melindungi properti intelektual hak cipta pada konten musik digital melawan pembajakan."

Hasil Token Dokumen1 :" pembajakan hak cipta terhadap konten musik digital masih menjadi masalah besar dalam industry musik. hal tersebut dikarenakan mudahnya proses pembajakan dan kemudahan distribusi konten digital melalui internet. isu perlindungan hak cipta menjadi hal sangat penting untuk diterapkan dalam industri musik. digital right management (drm) dan audio watermarking adalah cara yang bisa diterapkan untuk melindungi properti intelektual hak cipta pada konten musik digital melawan pembajakan."

Abstrak Dokumen2: “Telah dilakukan pembuatan suatu aplikasi kontrol switch menggunakan Silicon Controlled Rectifier (SCR) melalui parallel port. Menggunakan parallel port sebagai media antarmuka, SCR sebagai switch elektronis, optocoupler sebagai pengisolasi elektris dan bahasa pemograman yang digunakan Ezy Pascal. Dilakukan pengujian pada bagian output parallel port, bagian tegangan input, penggerak SCR dan bagian beban. Sistim yang dibuat membuktikan bahwa media parallel port dapat berfungsi sebagai media pengontrol SCR untuk mengaktifkan atau menonaktifkan beban.”

Hasil Token Dokumen2 : “telah dilakukan pembuatan suatu aplikasi kontrol switch menggunakan silicon controlled rectifier (scr) melalui parallel port. menggunakan parallel port sebagai media antarmuka, scr sebagai switch elektronis, optocoupler sebagai pengisolasi elektris dan bahasa pemograman yang digunakan ezy pascal. dilakukan pengujian pada bagian output parallel port, bagian tegangan input, penggerak scr dan bagian beban. sistim yang dibuat membuktikan bahwa media parallel port dapat berfungsi sebagai media pengontrol scr untuk mengaktifkan atau menonaktifkan beban.”

Abstrak Dokumen3: “Makalah ini mendeskripsikan model formal dari sistem interaktif berdasarkan sebuah studi kasus pada sistem kesehatan bergerak. Model ini mendeskripsikan sebuah session interaktif melibatkan professional medis yang menggunakan PDA untuk mengakses rekam medis pasien, dan terhubung pada server basis data dalam jaringan komunikasi nirkabel. Perilaku sistem dapat memunculkan masalah keselamatan bila dikaitkan dengan perilaku pemakai dalam pelayanan kesehatan. Akhirnya model sistem telah diverifikasi untuk menghindari masalah tersebut.”

Hasil Token Dokumen3: ”makalah ini mendeskripsikan model formal dari sistem interaktif berdasarkan sebuah studi kasus pada sistem kesehatan bergerak. model ini mendeskripsikan sebuah session interaktif melibatkan professional medis

yang menggunakan pda untuk mengakses rekam medis pasien, dan terhubung pada server basis data dalam jaringan komunikasi nirkabel. perilaku sistem dapat memunculkan masalah keselamatan bila dikaitkan dengan perilaku pemakai dalam pelayanan kesehatan. akhirnya model sistem telah diverifikasi untuk menghindari masalah tersebut."

Abstrak Dokumen4: "Sebuah simulator helikopter perlu dibuat agar benar-benar mirip dengan lingkungan cockpit sebenarnya termasuk instrumentasinya. Sebuah instrumen tersimulasi menggunakan teknologi mekatronika dimana sebuah sistem mekanik dikontrol secara elektronik untuk menunjukkan variabel terbang tertentu. Makalah ini menjelaskan proses rancang bangun sebuah indikator ketinggian tersimulasi yang digunakan pada sebuah simulator helikopter super puma. Indikator ini memiliki sebuah jarum utama yang dapat berputar 100 kali, sebuah jarum kecil yang dapat berputar 10 kali, dan sebuah jarum segi tiga yang dapat berputar sekali. Fitur keterbaruan makalah ini adalah bahwa jarum utama dan jarum kecil digerakkan oleh motor stepper sebagai pengganti syncro yang umumnya digunakan. Sebuah teknik kendali diusulkan untuk menggerakkan motor, dan kesalahan posisi diminimalkan memakai kompensator peranti lunak. Dari hasil eksperimen tingkat presisi berulang-ulang diperoleh kesimpulan bahwa indikator ketinggian tersimulasi yang dirancang bangun dapat bekerja dengan baik dengan tingkat ketelitian 0,03 (dial memory) untuk jarum utama dan 0,22 (dial memory) untuk jarum kecil."

Hasil Token Dokumen4: "sebuah simulator helikopter perlu dibuat agar benar-benar mirip dengan lingkungan cockpit sebenarnya termasuk instrumentasinya. sebuah instrumen tersimulasi menggunakan teknologi mekatronika dimana sebuah sistem mekanik dikontrol secara elektronik untuk menunjukkan variabel terbang tertentu. makalah ini menjelaskan proses rancang bangun sebuah indikator ketinggian tersimulasi yang digunakan pada sebuah simulator helikopter super puma. indikator ini memiliki sebuah jarum utama yang dapat berputar 100 kali, sebuah jarum kecil yang dapat berputar 10 kali, dan sebuah jarum segi tiga yang dapat berputar sekali.

fitur keterbaruan makalah ini adalah bahwa jarum utama dan jarum kecil digerakkan oleh motor stepper sebagai pengganti syncro yang umumnya digunakan. sebuah teknik kendali diusulkan untuk menggerakkan motor, dan kesalahan posisi diminimalkan memakai kompensator peranti lunak. dari hasil eksperimen tingkat presisi berulang-ulang diperoleh kesimpulan bahwa indikator ketinggian tersimulasi yang dirancang bangun dapat bekerja dengan baik dengan tingkat ketelitian 0,03 (dial memory) untuk jarum utama dan 0,22 (dial memory) untuk jarum kecil.“

Abstrak dokumen 5: “Kualitas layanan jaringan merupakan hal yang sangat penting untuk terus ditingkatkan. Oleh karena itu, perlu dilakukan layanan optimasi BTS yang bertujuan untuk pengoptimalan jaringan. Kegiatan optimasi jaringan dengan dilakukannya drive test. Begitu data telah terkumpul dari suatu area cakupan tertentu, data akan diproses lebih lanjut dengan software tool terpisah. Maintenance engineer dapat menggunakan data yang telah diproses untuk menganalisa unjuk kerja sistem di daerah tersebut. Layanan optimasi BTS diharapkan dapat mencapai target yang telah ditentukan untuk pengoptimalan seluruh sel disemua BTS.”

Hasil token Dokumen5 : “kualitas layanan jaringan merupakan hal yang sangat penting untuk terus ditingkatkan. oleh karena itu, perlu dilakukan layanan optimasi bts yang bertujuan untuk pengoptimalan jaringan. kegiatan optimasi jaringan dengan dilakukannya drive test. begitu data telah terkumpul dari suatu area cakupan tertentu, data akan diproses lebih lanjut dengan software tool terpisah. maintenance engineer dapat menggunakan data yang telah diproses untuk menganalisa unjuk kerja sistem di daerah tersebut. layanan optimasi bts diharapkan dapat mencapai target yang telah ditentukan untuk pengoptimalan seluruh sel disemua bts.”

2. ***Stopword***, merupakan tahapan penghapusan kata-kata yang tidak relevan dalam penentuan topik sebuah dokumen dan yang sering muncul pada sebuah dokumen

   Berikut wordlist yang digunakan dalam tugas akhir ini:

Ada; akan; antara; apa; atas; atau; bagi; baik; banyak; belum; demikian; kembali; berapa; bisa; buat; bukan; missal; secara; butuh; cukup; dalam; dan; dapat; dari; dengan; depan; di; dulu; tetapi; yaitu; bahkan; hanya; harus; ia; ini; itu; juga; kala; kami; jika; tahu; karena ke; kenapa; kendati; kira; ketika; kini; diri; mulai; kita; lagi; lain; lalu; lantas; mengerti; pilih; lebih; kata; maka; maksud; masing; kapan; masih; sementara; justru; mau; menjadi; mereka; namun; nanti; oleh; orang; pada; perlu; pernah; pula; seperti; sebagai; sedang; menjadi; agar; punya; memang; rendah; meski; terhadap; bahwa; saat; saja; bagai; sangat; saya; sejak; sejenak; semua; seusai; sekarang; menurut; meskipun; paling; siapa; sudah; tahun; tambah; mungkin; tanpa; mana; dahulu; telah; ternyata; tersebut; tidaktinggi; turun; untuk; walau; yang.

Hasil Stopword Dokumen1 :

Pembajakan; hak; cipta; konten; music; digital; masalah; industry; proses; distribusi; internet; isu; perlindungan; diterapkan; right; management; drm; audio; watermarking; diterapkan; property; intelektual; melawan.

Hasil Stopword Dokumen2 :

pembuatan; aplikasi; control; switch; silicon; controlled; rectifier; scr; parallel; port; parallel; port; media; antarmuka; scr; switch; elektronis; optocoupler; pengisolasi; elektris; bahasa; pemograman; ezy; pascal; pengujian; output; parallel; port; tegangan; input; penggerak; scr; beban; sistim; membuktikan; media; parallel; port; berfungsi; media; pengontrol; scr; mengaktifkan; menonaktifkan; beban.

Hasil Stopword Dokumen3 :

makalah; mendeskripsikan; model; formal; system; interaktif; studi; kasus; system; kesehatan; bergerak; model; mendeskripsikan; sebuah; session; interaktif; melibatkan; professional; medis; menggunakan; pda; mengakses; rekam; medis; pasien; terhubung; server; basis; data; jaringan; komunikasi; nirkabel; perilaku;

system; memunculkan; masalah; keselamatan; dikaitkan; perilaku; pemakai; pelayanan; kesehatan; model; system; diverifikasi; menghindari; masalah.

Hasil Stopword Dokumen4 :

simulator; helicopter; perlu; dibuat; mirip; lingkungan; cockpit; instrumentasinya; instrument; tersimulasi; menggunakan; teknologi; mekatronika; system; mekanik; dikontrol; elektronik; menunjukkan; variable; terbang; makalah; menjelaskan; proses; rancang; bangun; sebuah; indicator; ketinggian; tersimulasi; digunakan; sebuah; simulator; helicopter; super; puma; indicator; sebuah; jarum; utama; berputar; 100; kali; sebuah; jarum; kecil; berputar; 10; kali; sebuah; jarum; segitiga; berputar; sekali; fitur; keterbaruan; makalah; jarum; utama; jarum; kecil; digerakkan; motor; stepper; pengganti; syncro; umumnya; digunakan; sebuah; teknik; kendali; diusulkan; menggerakkan; motor; kesalahan; posisi; diminimalkan; kompensator; peranti; lunak; hasil; eksperimen; tingkat; presisi; berulangulang; indicator; ketinggian; tersimulasi; dirancang; bangun; bekerja; tingkat; ketelitian; 0,03; dialmemory; jarum; utama; 0,22; dialmemory; jarum; kecil.

Hasil Stopword Dokumen5 :

kualitas; layanan; jaringan; hal; ditingkatkan; dilakukan; layanan; optimasi; bts; pengoptimalan; jaringan; optimasi; jaringan; dilakukannya; drive; test; data; terkumpul; area; cakupan; data; diproses; software; tool; maintenance; engineer; data; diproses; menganalisa unjuk; kerja; system; daerah; layanan; optimasi; bts; mencapai; target; ditentukan; pengoptimalan; sel; disemua; bts.

3. *Stemming*, merupakan tahapan pengubahan suatu kata menjadi akar kata nya dengan menghilangkan imbuhan awal atau akhir pada kata tersebut

Hasil Stemming Dokumen1 : bajak; hak; cipta; konten; musik; digital; masalah; industri; proses; distribusi; internet; isu; lindung; terapkan; right; management; drm; audio; watermarking; property; intelektual; lawan.

Hasil Stemming Dokumen2 : buat; aplikasi; control; switch; silicon; rectifier; scr; port; parallel; media; antarmuka; elektronis; optocoupler; isolasi; elektris; bahasa; program; ezy; pascal; uji; output; tegang; input; gerak; beban; sistim; bukti; fungsi; aktif.

Hasil Stemming Dokumen3 : Makalah; deskripsi; model; formal; sistem; interaktif; studi; kasus; sehat; gerak; sebuah; session; libat; professional; medis; guna; pda; akses; rekam; pasien; hubung; server; basis; data; jaringan; komunikasi; nirkabel; perilaku; muncul; masalah; selamat; kait; pemakai; layan; verifikasi; hindari.

Hasil Stemming Dokumen4 : Simulator; helicopter; lingkungan; cockpit; instrument; simulasi; teknologi; mekatronika; sistem; mekanik; kontrol; elektronik; variable; terbang; proses; rancang; bangun; indikator; tinggi; super; puma; jarum; utama; kecil; segitiga; fitur; gerak; motor; stepper; syncro; teknik; kendali; posisi kompensator; peranti; lunak; eksperimen; tingkat; presisi; berulangulang; tinggi; dial; memory.

Hasil Stemming Dokumen5 : kualitas; layanan; jaringan; tingkat; optimasi; bts; optimal; drive; test; data; kumpul; area; proses; software; tool; maintenance; engineer; analisa; sistem; daerah; sel.

Berikut adalah tabel hasil dari preprosesing data untuk digunakan ke tahapan penentuan bobot setiap kata didalam sebuah dokumen.

Tabel III.1. Stemming Dokumen 1

Kata	Jumlah kata
bajak;	3
hak;	3
cipta;	3
konten;	3
musik;	4
digital;	4
masalah;	1
industri;	2
proses;	1
distribusi;	1
internet;	1

isu;	1
lindung;	2
terapkan;	2
right;	1
management;	1
drm;	1
audio;	1
watermarking;	1
property;	1
intelektual;	1
lawan.	1

ezy;	1
pascal;	1
uji;	1
output;	1
tegang;	1
input;	1
gerak;	1
beban;	2
sistim;	1
bukti;	1
fungsi;	1
aktif;	2

Tabel III.2. Stemming dokumen 2

Kata	Jumlah kata
Buat;	1
aplikasi;	1
control;	3
switch;	2
silicon;	1
rectifier;	1
scr;	4
parallel;	4
port;	4
media;	3
antarmuka;	1
elektronis;	1
optocoupler;	1
isolasi;	1
elektris;	1
bahasa;	1
program;	1

Tabel III.3. Stemming Dokumen 3

Kata	Jumlah kata
Makalah;	1
deskripsi;	2
model;	3
formal;	1
sistem;	4
interaktif;	2
studi;	1
kasus;	1
sehat;	2
gerak;	1
session;	1
libat;	1
professional;	1
medis;	2
guna;	1
pda;	1

akses;	1
rekam;	1
pasien;	1
hubung;	1
server;	1
basis;	1
data;	1
jaringan;	1
komunikasi;	1
nirkabel;	1
perilaku;	2
muncul;	1
masalah;	2
selamat;	1
kait;	1
pemakai;	1
layan;	1
verifikasi;	1
hindari.	1

Tabel III.4. Stemming Dokumen 4

Kata	Jumlah kata
Simulator;	2
helicopter;	2
lingkungan;	1
cockpit;	1
instrument;	2
simulasi;	3
teknologi;	1
mekatronika;	1
sistem;	1
mekanik;	1
kontrol;	1
elektronik;	1
variable;	1
terbang;	1
proses;	1
rancang;	2
bangun;	2
indikator;	3
super;	1
puma;	1
jarum;	7
utama;	3
kecil;	3
segitiga;	1
fitur;	1
gerak;	2
motor;	2
stepper;	1
Tinggi;	2
syncro;	1
teknik;	1
kendali;	1
posisi;	1
kompensator;	1
peranti;	1
lunak;	1
eksperimen;	1
presisi;	1
berulangulang;	1
dial;	2
memory.	2

Tabel III.5. Stemming Dokumen 5

Kata	Jumlah kata
Kualitas;	1
layanan;	3
jaringan;	3
tingkat;	1
optimasi;	3
bts;	3
optimal;	2
drive;	1
test;	1
data;	3
kumpul;	1
area;	1
proses;	2
software;	1
tool;	1
maintenance;	1
engineer;	1
analisa;	1
sistem;	1
daerah;	1
sel.	1

Setelah tahapan ini selesai hasil dari *stemming* digunakan sebagai data inputan untuk menghitung bobot setiap kata yang ada didokumen dengan menggunakan Pembobotan TF-IDF

III.1.4.2. Pembobotan TF-IDF

Pada tahap pembobotan digunakan sebuah rumus TF-IDF untuk menentukan tingkat similaritas antar dokumen. TF (*Term Frequency*) adalah frekuensi dari kemunculan sebuah *term* dalam dokumen yang bersangkutan. Oleh sebab itu, TF memiliki nilai yang bervariasi dari satu dokumen ke dokumen yang lain bergantung pada sebuah *term* dalam sebuah dokumen yang tersedia.

Sedangkan IDF (*Inverse Document Frequency*) adalah sebuah statistik global yang mengkarakteristikkan sebuah *term* dalam keseluruhan korelasi dokumen. IDF merupakan sebuah perhitungan dari bagaimana *term* didistribusikan secara luas pada kalimat-kalimat.

Perhitungan bobot dari *term* tertentu dalam sebuah kalimat dengan menggunakan TF-IDF menunjukkan bahwa kalimat yang dianggap similaritas adalah *term* yang banyak muncul dalam suatu kalimat.

Berikut penerapan TF-IDF pada karya ilmiah ini. Sebagai sampel, terdapat 5 buah dokumen yang sudah melewati tahap preprosesing:

Dokumen1 : bajak; hak; cipta; konten; musik; digital; masalah; industri; proses; distribusi; internet; isu; lindung; terapkan; right; management; drm; audio; watermarking; property; intelektual; lawan.

Dokumen2 : buat; aplikasi; control; switch; silicon; rectifier; scr; port; parallel; media; antarmuka; elektronis; optocoupler; isolasi; elektris; bahasa; program; ezy; pascal; uji; output; tegang; input; gerak; beban; sistim; bukti; fungsi; aktif; beban.

Dokumen3 : Makalah; deskripsi; model; formal; sistem; interaktif; studi; kasus; sehat; gerak; sebuah; session; libat; professional; medis; guna; pda; akses; rekam; pasien; hubung; server; basis; data; jaringan; komunikasi; nirkabel; perilaku; muncul; masalah; selamat; kait; pemakai; layan; verifikasi; hindari.

Dokumen4 : Simulator; helicopter; lingkungan; cockpit; instrument; simulasi; teknologi; mekatronika; sistem; mekanik; kontrol; elektronik; variable; terbang; proses; rancang; bangun; indikator; tinggi; super; puma; jarum; utama; kecil; segitiga; fitur; gerak; motor; stepper; syncro; teknik; kendali; posisi kompensator; peranti; lunak; eksperimen; tingkat; presisi; berulangulang; tinggi; dial; memory.

Dokumen5 : kualitas; layanan; jaringan; tingkat; optimasi; bts; optimal; drive; test; data; kumpul; area; proses; software; tool; maintenance; engineer; analisa; sistem; daerah; sel.

Berdasarkan 5 dokumen diatas, maka dilakukan proses pembangunan indeks untuk membentuk *lexicon* (kamus token) dan pembobotan dengan TF-IDF hingga didapat nilai w(bobot) sesuai dengan rumus yang tersedia.

Berikut ini diberikan *lexicon* yang terbentuk dan perhitungan untuk token “motor”. Token “motor” dimiliki oleh dokumen 1 dan 3 sebanyak 1. Jadi df untuk token “motor” adalah 2. Berikut perhitungan IDF untuk token “motor”

$$IDF\ (sistem) = \log(\frac{jumlah\ dokumen}{jumlah\ dokumen\ yang\ memuat\ "sistem"})$$

$$IDF_{(sistem)} = log\left(\frac{4}{7}\right)$$

$IDF_{(sistem)}$ =-0.146128036

Berikut adalah perhitungan TF-IDF untuk w dari token “sistem”:

$$w_{(sistem)} = TF_{sistem} * IDF_{(sistem)}$$

$$w_{(sistem)} = 4 * -0.146128036$$

$w_{(sistem)}$ = -0.584512143

Langkah diatas dilakukan terhadap semua token yang dihasilkan. Maka dihasilkan data seperti table dibawah ini:

Tabel III.6. Perhitungan TF

Token	TF					TOTAL
	D1	D2	**D3**	**D4**	**D5**	
bajak	3					3
hak	3					3
cipta	3					3
konten	3					3

musik	4					4
digital	4					4
masalah	1		2			3
industri	2					2
proses	1			1	2	4
distribusi	1					1
internet	1					1
isu	1					1
lindung	2					2
terapkan	2					2
right	1					1
management	1					1
drm	1					1
audio	1					1
watermarking	1					1
property	1					1
intelektual	1					1
lawan	1					1
Buat		1				1
aplikasi		1				1
kontrol		3		1		4

switch		2				2
silicon		1				1
rectifier		1				1
scr		4				4
parallel		4				4
port		4				4
media		3				3
antarmuka		1				1
elektronis		1				1
optocoupler		1				1
isolasi		1				1
elektris		1				1
bahasa		1				1
program		1				1
ezy		1				1
pascal		1				1
uji		1				1
output		1				1
tegang		1				1
input		1				1
gerak		1	1	2		4

beban		2				2
sistem		1	4	1	1	7
bukti		1				1
fungsi		1				1
aktif		2				2
Makalah			1			1
deskripsi			2			2
model			3			3
formal			1			1
interaktif			2			2
studi			1			1
kasus			1			1
sehat			2			2
session			1			1
libat			1			1
professional			1			1
medis			2			2
guna			1			1
pda			1			1
akses			1			1
rekam			1			1

pasien			1			1
hubung			1			1
server			1			1
basis			1			1
data			1		3	4
jaringan			1		3	4
komunikasi			1			1
nirkabel			1			1
perilaku			2			2
muncul			1			1
selamat			1			1
kait			1			1
pemakai			1			1
layan			1		3	4
verifikasi			1			1
hindari			1			1
Simulator				2		2
helicopter				2		2
lingkungan				1		1
cockpit				1		1
instrument				2		2

simulasi				3		3
teknologi				1		1
mekatronika				1		1
mekanik				1		1
elektronik				1		1
variable				1		1
terbang				1		1
rancang				2		2
bangun				2		2
indikator				3		3
super				1		1
puma				1		1
jarum				7		7
utama				3		3
kecil				3		3
segitiga				1		1
fitur				1		1
motor				2		2
stepper				1		1
Tinggi				2		2
syncro				1		1

teknik				1		1
kendali				1		1
posisi				1		1
kompensator				1		1
peranti				1		1
lunak				1		1
eksperimen				1		1
presisi				1		1
berulangulang				1		1
dial				2		2
memory				2		2
Kualitas					1	1
tingkat					1	1
optimasi					3	3
bts					3	3
optimal					2	2
drive					1	1
test					1	1
kumpul					1	1
area					1	1
software					1	1

tool					1	1
maintenance					1	1
engineer					1	1
analisa					1	1
daerah					1	1
sel					1	1

Tabel III.7. Perhitungan IDF

Token	Total kata	IDF
bajak	3	0.22184875
hak	3	0.22184875
cipta	3	0.22184875
konten	3	0.22184875
musik	4	0.09691001
digital	4	0.09691001
masalah	3	0.22184875
industri	2	0.39794001
proses	4	0.09691001
distribusi	1	0.69897
internet	1	0.69897
isu	1	0.69897
lindung	2	0.39794001
terapkan	2	0.39794001
right	1	0.69897
management	1	0.69897
drm	1	0.69897
audio	1	0.69897

watermarking	1	0.69897
property	1	0.69897
intelektual	1	0.69897
lawan	1	0.69897
Buat	1	0.69897
aplikasi	1	0.69897
kontrol	4	0.09691001
switch	2	0.39794001
silicon	1	0.69897
rectifier	1	0.69897
scr	4	0.09691001
parallel	4	0.09691001
port	4	0.09691001
media	3	0.22184875
antarmuka	1	0.69897
elektronis	1	0.69897
optocoupler	1	0.69897
isolasi	1	0.69897
elektris	1	0.69897
bahasa	1	0.69897
program	1	0.69897
ezy	1	0.69897
pascal	1	0.69897
uji	1	0.69897
output	1	0.69897
tegang	1	0.69897
input	1	0.69897
gerak	4	0.09691001
beban	2	0.39794001
sistem	7	-0.146128
bukti	1	0.69897
fungsi	1	0.69897
aktif	2	0.39794001

Makalah	1	0.69897
deskripsi	2	0.39794001
model	3	0.22184875
formal	1	0.69897
interaktif	2	0.39794001
studi	1	0.69897
kasus	1	0.69897
sehat	2	0.39794001
session	1	0.69897
libat	1	0.69897
professional	1	0.69897
medis	2	0.39794001
guna	1	0.69897
pda	1	0.69897
akses	1	0.69897
rekam	1	0.69897
pasien	1	0.69897
hubung	1	0.69897
server	1	0.69897
basis	1	0.69897
data	4	0.09691001
jaringan	4	0.09691001
komunikasi	1	0.69897
nirkabel	1	0.69897
perilaku	2	0.39794001
muncul	1	0.69897
selamat	1	0.69897
kait	1	0.69897
pemakai	1	0.69897
layan	4	0.09691001
verifikasi	1	0.69897
hindari	1	0.69897
Simulator	2	0.39794001

helicopter	2	0.39794001
lingkungan	1	0.69897
cockpit	1	0.69897
instrument	2	0.39794001
simulasi	3	0.22184875
teknologi	1	0.69897
mekatronika	1	0.69897
mekanik	1	0.69897
elektronik	1	0.69897
variable	1	0.69897
terbang	1	0.69897
rancang	2	0.39794001
bangun	2	0.39794001
indikator	3	0.22184875
super	1	0.69897
puma	1	0.69897
jarum	7	-0.146128
utama	3	0.22184875
kecil	3	0.22184875
segitiga	1	0.69897
fitur	1	0.69897
motor	2	0.39794001
stepper	1	0.69897
Tinggi	2	0.39794001
syncro	1	0.69897
teknik	1	0.69897
kendali	1	0.69897
posisi	1	0.69897
kompensator	1	0.69897
peranti	1	0.69897
lunak	1	0.69897
eksperimen	1	0.69897
presisi	1	0.69897

berulangulang	1	0.69897
dial	2	0.39794001
memory	2	0.39794001
Kualitas	1	0.69897
tingkat	1	0.69897
optimasi	3	0.22184875
bts	3	0.22184875
optimal	2	0.39794001
drive	1	0.69897
test	1	0.69897
kumpul	1	0.69897
area	1	0.69897
software	1	0.69897
tool	1	0.69897
maintenance	1	0.69897
engineer	1	0.69897
analisa	1	0.69897
daerah	1	0.69897
sel	1	0.69897

Tabel III.9. Perhitungan Bobot setiap Token

Term	Token	idf				
		D1	**D2**	**D3**	**D4**	**D5**
bajak	0.22184875	0.665546249	0	0	0	0
hak	0.22184875	0.665546249	0	0	0	0
cipta	0.22184875	0.665546249	0	0	0	0
konten	0.22184875	0.665546249	0	0	0	0
musik	0.096910013	0.387640052	0	0	0	0
digital	0.096910013	0.387640052	0	0	0	0
masalah	0.22184875	0.22184875	0	0.443697499	0	0
industri	0.397940009	0.795880017	0	0	0	0
proses	0.096910013	0.096910013	0	0	0.096910013	0.193820026
distribusi	0.698970004	0.698970004	0	0	0	0
internet	0.698970004	0.698970004	0	0	0	0

isu	0.698970004	0.698970004	0	0	0	0
lindung	0.397940009	0.795880017	0	0	0	0
terapkan	0.397940009	0.795880017	0	0	0	0
right	0.698970004	0.698970004	0	0	0	0
management	0.698970004	0.698970004	0	0	0	0
drm	0.698970004	0.698970004	0	0	0	0
audio	0.698970004	0.698970004	0	0	0	0
watermarking	0.698970004	0.698970004	0	0	0	0
property	0.698970004	0.698970004	0	0	0	0
intelektual	0.698970004	0.698970004	0	0	0	0
lawan	0.698970004	0.698970004	0	0	0	0
Buat	0.698970004	0	0.698970004	0	0	0
aplikasi	0.698970004	0	0.698970004	0	0	0

kontrol	0.096910013	0	0.290730039	0	0.096910013	0
switch	0.397940009	0	0.795880017	0	0	0
silicon	0.698970004	0	0.698970004	0	0	0
rectifier	0.698970004	0	0.698970004	0	0	0
scr	0.096910013	0	0.387640052	0	0	0
parallel	0.096910013	0	0.387640052	0	0	0
port	0.096910013	0	0.387640052	0	0	0
media	0.22184875	0	0.665546249	0	0	0
antarmuka	0.698970004	0	0.698970004	0	0	0
elektronis	0.698970004	0	0.698970004	0	0	0
optocoupler	0.698970004	0	0.698970004	0	0	0
isolasi	0.698970004	0	0.698970004	0	0	0
elektris	0.698970004	0	0.698970004	0	0	0
bahasa	0.698970004	0	0.698970004	0	0	0

program	0.698970004	0	0.698970004	0	0	0
ezy	0.698970004	0	0.698970004	0	0	0
pascal	0.698970004	0	0.698970004	0	0	0
uji	0.698970004	0	0.698970004	0	0	0
output	0.698970004	0	0.698970004	0	0	0
tegang	0.698970004	0	0.698970004	0	0	0
input	0.698970004	0	0.698970004	0	0	0
gerak	0.096910013	0	0.096910013	0.096910013	0.193820026	0
beban	0.397940009	0	0.795880017	0	0	0
sistem	-0.146128036	0	-0.146128036	-0.584512143	-0.146128036	-0.146128036
bukti	0.698970004	0	0.698970004	0	0	0
fungsi	0.698970004	0	0.698970004	0	0	0
aktif	0.397940009	0	0.795880017	0	0	0
Makalah	0.698970004	0	0	0.698970004	0	0

deskripsi	0.397940009	0	0	0.795880017	0	0
model	0.22184875	0	0	0.665546249	0	0
formal	0.698970004	0	0	0.698970004	0	0
interaktif	0.397940009	0	0	0.795880017	0	0
studi	0.698970004	0	0	0.698970004	0	0
kasus	0.698970004	0	0	0.698970004	0	0
sehat	0.397940009	0	0	0.795880017	0	0
session	0.698970004	0	0	0.698970004	0	0
libat	0.698970004	0	0	0.698970004	0	0
professional	0.698970004	0	0	0.698970004	0	0
medis	0.397940009	0	0	0.795880017	0	0
guna	0.698970004	0	0	0.698970004	0	0
pda	0.698970004	0	0	0.698970004	0	0
akses	0.698970004	0	0	0.698970004	0	0

rekam	0.698970004	0	0	0.698970004	0	0
pasien	0.698970004	0	0	0.698970004	0	0
hubung	0.698970004	0	0	0.698970004	0	0
server	0.698970004	0	0	0.698970004	0	0
basis	0.698970004	0	0	0.698970004	0	0
data	0.096910013	0	0	0.096910013	0	0.290730039
jaringan	0.096910013	0	0	0.096910013	0	0.290730039
komunikasi	0.698970004	0	0	0.698970004	0	0
nirkabel	0.698970004	0	0	0.698970004	0	0
perilaku	0.397940009	0	0	0.795880017	0	0
muncul	0.698970004	0	0	0.698970004	0	0
selamat	0.698970004	0	0	0.698970004	0	0
kait	0.698970004	0	0	0.698970004	0	0
pemakai	0.698970004	0	0	0.698970004	0	0

layan	0.096910013	0	0	0.096910013	0	0.290730039
verifikasi	0.698970004	0	0	0.698970004	0	0
hindari	0.698970004	0	0	0.698970004	0	0
Simulator	0.397940009	0	0	0	0.795880017	0
helicopter	0.397940009	0	0	0	0.795880017	0
lingkungan	0.698970004	0	0	0	0.698970004	0
cockpit	0.698970004	0	0	0	0.698970004	0
instrument	0.397940009	0	0	0	0.795880017	0
simulasi	0.22184875	0	0	0	0.665546249	0
teknologi	0.698970004	0	0	0	0.698970004	0
mekatronika	0.698970004	0	0	0	0.698970004	0
mekanik	0.698970004	0	0	0	0.698970004	0
elektronik	0.698970004	0	0	0	0.698970004	0
variable	0.698970004	0	0	0	0.698970004	0

terbang	0.698970004	0	0	0	0.698970004	0
rancang	0.397940009	0	0	0	0.795880017	0
bangun	0.397940009	0	0	0	0.795880017	0
indikator	0.22184875	0	0	0	0.665546249	0
super	0.698970004	0	0	0	0.698970004	0
puma	0.698970004	0	0	0	0.698970004	0
jarum	-0.146128036	0	0	0	-1.02289625	0
utama	0.22184875	0	0	0	0.665546249	0
kecil	0.22184875	0	0	0	0.665546249	0
segitiga	0.698970004	0	0	0	0.698970004	0
fitur	0.698970004	0	0	0	0.698970004	0
motor	0.397940009	0	0	0	0.795880017	0
stepper	0.698970004	0	0	0	0.698970004	0
Tinggi	0.397940009	0	0	0	0.795880017	0

syncro	0.698970004	0	0	0	0.698970004	0
teknik	0.698970004	0	0	0	0.698970004	0
kendali	0.698970004	0	0	0	0.698970004	0
posisi	0.698970004	0	0	0	0.698970004	0
kompensator	0.698970004	0	0	0	0.698970004	0
peranti	0.698970004	0	0	0	0.698970004	0
lunak	0.698970004	0	0	0	0.698970004	0
eksperimen	0.698970004	0	0	0	0.698970004	0
presisi	0.698970004	0	0	0	0.698970004	0
berulangulang	0.698970004	0	0	0	0.698970004	0
dial	0.397940009	0	0	0	0.795880017	0
memory	0.397940009	0	0	0	0.795880017	0
Kualitas	0.698970004	0	0	0	0	0.698970004
tingkat	0.698970004	0	0	0	0	0.698970004

optimasi	0.22184875	0	0	0	0	0.665546249
bts	0.22184875	0	0	0	0	0.665546249
optimal	0.397940009	0	0	0	0	0.795880017
drive	0.698970004	0	0	0	0	0.698970004
test	0.698970004	0	0	0	0	0.698970004
kumpul	0.698970004	0	0	0	0	0.698970004
area	0.698970004	0	0	0	0	0.698970004
software	0.698970004	0	0	0	0	0.698970004
tool	0.698970004	0	0	0	0	0.698970004
maintenance	0.698970004	0	0	0	0	0.698970004
engineer	0.698970004	0	0	0	0	0.698970004
analisa	0.698970004	0	0	0	0	0.698970004
daerah	0.698970004	0	0	0	0	0.698970004
sel	0.698970004	0	0	0	0	0.698970004

III.1.4.3. Analisis clustering

Setelah melakukan data *preprocessing* dan perhitungan bobot disetiap kata pada kalimat, Tahapan selanjutnya adalah Analisis *clustering* dengan menggunakan algoritma K-Means. Adapun tahapannya adalah sebagai berikut :

Algoritma dari metode k-Means itu sendiri adalah sebagai berikut:

1. Dimulai dengan menentukan jumlah cluster. Pada kasus ini akan dibentuk 2 cluster untuk membagi data yang ada pada tabel III.8.
2. Menentukan nilai centroid secara acak. Berikut ini nilai centroid awal yang akan digunakan:

Tabel III.10. Nilai Centroid secara acak

centroid awal		
Token	**Centroid 1**	**Centroid 2**
bajak	0	0
hak	0	0
cipta	0	0
konten	0	0
musik	0	0
digital	0	0
masalah	0.4436975	0
industri	0	0
proses	0	0.19382003
distribusi	0	0

internet	0	0
isu	0	0
lindung	0	0
terapkan	0	0
right	0	0
management	0	0
drm	0	0
audio	0	0
watermarking	0	0
property	0	0
intelektual	0	0
lawan	0	0
Buat	0	0
aplikasi	0	0
kontrol	0	0
switch	0	0
silicon	0	0
rectifier	0	0
scr	0	0
parallel	0	0
port	0	0

media	0	0
antarmuka	0	0
elektronis	0	0
optocoupler	0	0
isolasi	0	0
elektris	0	0
bahasa	0	0
program	0	0
ezy	0	0
pascal	0	0
uji	0	0
output	0	0
tegang	0	0
input	0	0
gerak	0.09691001	0
beban	0	0
sistem	-0.5845121	-0.146128
bukti	0	0
fungsi	0	0
aktif	0	0
Makalah	0.69897	0

deskripsi	0.79588002	0
model	0.66554625	0
formal	0.69897	0
interaktif	0.79588002	0
studi	0.69897	0
kasus	0.69897	0
sehat	0.79588002	0
session	0.69897	0
libat	0.69897	0
professional	0.69897	0
medis	0.79588002	0
guna	0.69897	0
pda	0.69897	0
akses	0.69897	0
rekam	0.69897	0
pasien	0.69897	0
hubung	0.69897	0
server	0.69897	0
basis	0.69897	0
data	0.09691001	0.29073004
jaringan	0.09691001	0.29073004

komunikasi	0.69897	0
nirkabel	0.69897	0
perilaku	0.79588002	0
muncul	0.69897	0
selamat	0.69897	0
kait	0.69897	0
pemakai	0.69897	0
layan	0.09691001	0.29073004
verifikasi	0.69897	0
hindari	0.69897	0
Simulator	0	0
helicopter	0	0
lingkungan	0	0
cockpit	0	0
instrument	0	0
simulasi	0	0
teknologi	0	0
mekatronika	0	0
mekanik	0	0
elektronik	0	0
variable	0	0

terbang	0	0
rancang	0	0
bangun	0	0
indikator	0	0
super	0	0
puma	0	0
jarum	0	0
utama	0	0
kecil	0	0
segitiga	0	0
fitur	0	0
motor	0	0
stepper	0	0
Tinggi	0	0
syncro	0	0
teknik	0	0
kendali	0	0
posisi	0	0
kompensator	0	0
peranti	0	0
lunak	0	0

eksperimen	0	0
presisi	0	0
berulangulang	0	0
dial	0	0
memory	0	0
Kualitas	0	0.69897
tingkat	0	0.69897
optimasi	0	0.66554625
bts	0	0.66554625
optimal	0	0.79588002
drive	0	0.69897
test	0	0.69897
kumpul	0	0.69897
area	0	0.69897
software	0	0.69897
tool	0	0.69897
maintenance	0	0.69897
engineer	0	0.69897
analisa	0	0.69897
daerah	0	0.69897
sel	0	0.69897

3. Setelah selesai menentukan nilai centroid awal, tahap selanjutnya adalah melakukan penghitungan jarak setiap data terhadap centroid awal yang telah dibentuk menggunakan rumus euclidiance distance. Rumus ini akan menentukan penempatan setiap data pada cluster tertentu.. Penghitungan jarak antara data centroid awal terhadap masing-masing cluster menggunakn rumus (II.1)

Menghitung nilai centroid yang paling dekat dari setiap dokumen menggunakan euclidien distance

1) Menghitung jarak antara Dokumen Pertama (D1) ke centroid Pertama (C1)

$$d_{D1.Cent1} = \sqrt{\left(|X_{i1} - X_{i1}|^2 + |X_{i2} - X_{i2}|^2 + .. + \left|X_{ip} - X_{ip}\right|^2\right)}$$

$$d_{D1.Cent1} = \sqrt{\mathbf{24.63152601}}$$

=**4.963015818**

2) Menghitung jarak antara Dokumen Pertama (D1) ke centroid Pertama (C2)

$$d_{D1.Cent2} = \sqrt{\mathbf{17.55089177}}$$

=**4.189378446**

3) Menghitung jarak antara Dokumen Pertama (D2) ke centroid Pertama (C1)

$$d_{D2.Cent1} = \sqrt{\mathbf{27.42532556}}$$

=**5.236919472**

4) Menghitung jarak antara Dokumen Pertama (D2) ke centroid Pertama (C2)

$$d_{D2.Cent2} = \sqrt{\mathbf{20.33229357}}$$

Tabel III.11. =4.509134459

5) Menghitung jarak antara Dokumen Pertama (D3) ke centroid Pertama (C1)

$$d_{D3.Cent1} = \sqrt{\mathbf{0}}$$

Tabel III.12. =0

6) Menghitung jarak antara Dokumen Pertama (D3) ke centroid Pertama (C2)

$$d_{D3.Cent2} = \sqrt{\mathbf{23.26623629}}$$

Tabel III.13. =4.823508711

7) Menghitung jarak antara Dokumen Pertama (D4) ke centroid Pertama (C1)

$$d_{D4.Cent1} = \sqrt{\mathbf{35.04814026}}$$

Tabel III.14. = 5.92014698

8) Menghitung jarak antara Dokumen Pertama (D4) ke centroid Pertama (C2)

$$d_{D4.Cent2} = \sqrt{\mathbf{27.93632517}}$$

Tabel III.15. = 5.285482492

9) Menghitung jarak antara Dokumen Pertama (D5) ke centroid Pertama (C1)

$$d_{D5.Cent1} = \sqrt{\mathbf{23.26623629}}$$

Tabel III.16. = 4.823508711

10) Menghitung jarak antara Dokumen Pertama (D5) ke centroid Pertama (C2)

$$d_{D5.Cent2} = \sqrt{\mathbf{0}} \quad =0$$

Hasil dari seluruh perhitungan di lampirkan pada Tabel III.10

Tabel III.10 Euclidean Distance

D1,C1	**D1,C2**	**D2,C1**	**D2,C2**	**D3,C1**	**D3,C2**	**D4,C1**	**D4,C2**	**D5,C1**	**D5,C2**
0.44295181	0.44295181	0	0	0	0	0	0	0	0
0.44295181	0.44295181	0	0	0	0	0	0	0	0
0.44295181	0.44295181	0	0	0	0	0	0	0	0
0.44295181	0.44295181	0	0	0	0	0	0	0	0
0.15026481	0.15026481	0	0	0	0	0	0	0	0
0.15026481	0.15026481	0	0	0	0	0	0	0	0
0.04921686 7	0.04921686 8	0.19686747 1	0	0	0.19686747 1	0.19686747 1	0	0.19686747 1	0

0.633425001	0.633425001	0	0	0	0	0	0	0	0
0.009391551	0.009391551	0	0.037566202	0	0.037566202	0.009391551	0.009391551	0.037566202	0
0.488559066	0.488559066	0	0	0	0	0	0	0	0
0.488559066	0.488559066	0	0	0	0	0	0	0	0
0.488559066	0.488559066	0	0	0	0	0	0	0	0
0.633425001	0.633425001	0	0	0	0	0	0	0	0
0.633425001	0.633425001	0	0	0	0	0	0	0	0
0.488559066	0.488559066	0	0	0	0	0	0	0	0
0.488559066	0.488559066	0	0	0	0	0	0	0	0
0.488559	0.488559	0	0	0	0	0	0	0	0

66	66								
0.488559066	0.488559066	0	0	0	0	0	0	0	0
0.488559066	0.488559066	0	0	0	0	0	0	0	0
0.488559066	0.488559066	0	0	0	0	0	0	0	0
0.488559066	0.488559066	0	0	0	0	0	0	0	0
0.488559066	0.488559066	0	0	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.084523956	0.084523956	0	0	0.009391551	0.009391551	0	0
0	0	0.633425001	0.633425001	0	0	0	0	0	0

0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.15026481	0.15026481	0	0	0	0	0	0
0	0	0.15026481	0.15026481	0	0	0	0	0	0
0	0	0.15026481	0.15026481	0	0	0	0	0	0
0	0	0.44295181	0.44295181	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.4885590	0.4885590	0	0	0	0	0	0

		66	66						
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0

0.009391551	0	0	0.009391551	0	0.009391551	0.009391551	0.037566202	0.009391551	0
0	0	0.633425001	0.633425001	0	0	0	0	0	0
0.341654445	0.021353403	0.192180625	0	0	0.192180625	0.192180625	0	0.192180625	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.488559066	0.488559066	0	0	0	0	0	0
0	0	0.633425001	0.633425001	0	0	0	0	0	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.633425001	0	0.633425001	0	0	0.633425001	0.633425001	0	0.633425001	0
0.44295181	0	0.44295181	0	0	0.44295181	0.44295181	0	0.44295181	0
0.4885590	0	0.4885590	0	0	0.4885590	0.4885590	0	0.4885590	0

66		66			66	66		66	
0.633425001	0	0.633425001	0	0	0.633425001	0.633425001	0	0.633425001	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.633425001	0	0.633425001	0	0	0.633425001	0.633425001	0	0.633425001	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.633425001	0	0.633425001	0	0	0.633425001	0.633425001	0	0.633425001	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0

0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.009391551	0.084523956	0.009391551	0.084523956	0	0.037566202	0.009391551	0.084523956	0.037566202	0
0.009391551	0.084523956	0.009391551	0.084523956	0	0.037566202	0.009391551	0.084523956	0.037566202	0
0.488559	0	0.488559	0	0	0.488559	0.488559	0	0.488559	0

66		66			66	66		66	
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.633425001	0	0.633425001	0	0	0.633425001	0.633425001	0	0.633425001	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.009391551	0.084523956	0.009391551	0.084523956	0	0.037566202	0.009391551	0.084523956	0.037566202	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0
0.488559066	0	0.488559066	0	0	0.488559066	0.488559066	0	0.488559066	0

0	0	0	0	0	0	0.633425001	0.633425001	0	0
0	0	0	0	0	0	0.633425001	0.633425001	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.633425001	0.633425001	0	0
0	0	0	0	0	0	0.44295181	0.44295181	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.4885590	0.4885590	0	0

						66	66		
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.633425001	0.633425001	0	0
0	0	0	0	0	0	0.633425001	0.633425001	0	0
0	0	0	0	0	0	0.4429518 1	0.4429518 1	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	1.046316738	1.046316738	0	0
0	0	0	0	0	0	0.4429518 1	0.4429518 1	0	0

0	0	0	0	0	0	0.44295181	0.44295181	0	0
0	0	0	0	0	0	0.48855906	0.48855906	0	0
0	0	0	0	0	0	0.48855906	0.48855906	0	0
0	0	0	0	0	0	0.63342501	0.63342501	0	0
0	0	0	0	0	0	0.48855906	0.48855906	0	0
0	0	0	0	0	0	0.63342501	0.63342501	0	0
0	0	0	0	0	0	0.48855906	0.48855906	0	0
0	0	0	0	0	0	0.48855906	0.48855906	0	0
0	0	0	0	0	0	0.48855906	0.48855906	0	0
0	0	0	0	0	0	0.4885590	0.4885590	0	0

						66	66		
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.488559066	0.488559066	0	0
0	0	0	0	0	0	0.633425001	0.633425001	0	0
0	0	0	0	0	0	0.633425001	0.633425001	0	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0

0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.44295181	0	0.44295181	0	0.44295181	0	0.44295181	0.44295181	0
0	0.44295181	0	0.44295181	0	0.44295181	0	0.44295181	0.44295181	0
0	0.633425001	0	0.633425001	0	0.633425001	0	0.633425001	0.633425001	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.488559	0	0.488559	0	0.488559	0	0.488559	0.488559	0

	66		66		66		66	66	
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
0	0.488559066	0	0.488559066	0	0.488559066	0	0.488559066	0.488559066	0
24.63152601	**17.55089177**	**27.42532556**	**20.33229357**	**0**	**23.26623629**	**35.04814026**	**27.93632517**	**23.26623629**	**0**
4.963015818	**4.189378446**	**5.236919472**	**4.509134459**	**0**	**4.823508711**	**5.920146978**	**5.285482492**	**4.823508711**	**0**

Tabel III.11 Perhitungan jarak dokumen dengan centroid pada iterasi pertama

Dokumen	Centroid	Jarak
D(1)	Centroid 1	4.963015818
D(1)	**Centroid 2**	**4.189378446**
D(2)	Centroid 1	5.236919472
D(2)	**Centroid 2**	**4.509134459**
D(3)	**Centroid 1**	**0**
D(3)	Centroid 2	4.823508711
D(4)	Centroid 1	5.92014698
D(4)	**Centroid 2**	**5.285482492**
D(5)	Centroid 1	4.823508711
D(5)	**Centroid 2**	**0**

Tabel III.12pembagian dokumen kedalam cluster

Dokumen	Jarak	Masuk ke cluster
D(1)	**4.189378446**	2
D(2)	**4.509134459**	2
D(3)	0	1
D(4)	**5.285482492**	2
D(5)	0	2

4. Kemudian dilakukan perhitungan untuk menentukan posisi centroid yang baru. Dimana dihitung melalui rata-rata dari masing-masing vector yang merupakan anggota dari cluster. Didapatkan :

Tabel III.13 Centroid Rata Rata

Tabel perhitungan centroid C1 yang baru	rata rata cluster 1
0	0
0	0
0	0
0	0
0	0
0	0
0.443697499	0.443697499
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0

0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0.096910013	0.096910013
0	0
-0.584512143	-0.584512143
0	0
0	0
0	0
0.698970004	0.698970004
0.795880017	0.795880017
0.665546249	0.665546249
0.698970004	0.698970004
0.795880017	0.795880017
0.698970004	0.698970004
0.698970004	0.698970004
0.795880017	0.795880017
0.698970004	0.698970004

0.698970004	0.698970004
0.698970004	0.698970004
0.795880017	0.795880017
0.698970004	0.698970004
0.698970004	0.698970004
0.698970004	0.698970004
0.698970004	0.698970004
0.698970004	0.698970004
0.698970004	0.698970004
0.698970004	0.698970004
0.698970004	0.698970004
0.096910013	0.096910013
0.096910013	0.096910013
0.698970004	0.698970004
0.698970004	0.698970004
0.795880017	0.795880017
0.698970004	0.698970004
0.698970004	0.698970004
0.698970004	0.698970004
0.698970004	0.698970004
0.096910013	0.096910013
0.698970004	0.698970004
0.698970004	0.698970004
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0

0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0

0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0

Tabel III.14 Perhitungan Centroid C2 yang Baru

D1	**D2**	**D4**	**D5**	rata rata cluster 2
0.665546249	0	0	0	0.16638656
0.665546249	0	0	0	0.16638656
0.665546249	0	0	0	0.16638656
0.665546249	0	0	0	0.16638656
0.387640052	0	0	0	0.09691001
0.387640052	0	0	0	0.09691001
0.22184875	0	0	0	0.05546219
0.795880017	0	0	0	0.19897
0.096910013	0	0.096910013	0.193820026	0.09691001
0.698970004	0	0	0	0.1747425
0.698970004	0	0	0	0.1747425
0.698970004	0	0	0	0.1747425
0.795880017	0	0	0	0.19897
0.795880017	0	0	0	0.19897
0.698970004	0	0	0	0.1747425
0.698970004	0	0	0	0.1747425
0.698970004	0	0	0	0.1747425

0.698970004	0	0	0	0.1747425
0.698970004	0	0	0	0.1747425
0.698970004	0	0	0	0.1747425
0.698970004	0	0	0	0.1747425
0.698970004	0	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.290730039	0.096910013	0	0.09691001
0	0.795880017	0	0	0.19897
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.387640052	0	0	0.09691001
0	0.387640052	0	0	0.09691001
0	0.387640052	0	0	0.09691001
0	0.665546249	0	0	0.16638656
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425
0	0.096910013	0.193820026	0	0.07268251
0	0.795880017	0	0	0.19897
0	-0.14612804	-0.146128036	-0.146128036	-0.109596
0	0.698970004	0	0	0.1747425
0	0.698970004	0	0	0.1747425

0	0.795880017	0	0	0.19897
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0.290730039	0.07268251
0	0	0	0.290730039	0.07268251
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0.290730039	0.07268251
0	0	0	0	0
0	0	0	0	0

0	0	0.795880017	0	0.19897
0	0	0.795880017	0	0.19897
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.795880017	0	0.19897
0	0	0.665546249	0	0.16638656
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.795880017	0	0.19897
0	0	0.795880017	0	0.19897
0	0	0.665546249	0	0.16638656
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	-1.02289625	0	-0.2557241
0	0	0.665546249	0	0.16638656
0	0	0.665546249	0	0.16638656
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.795880017	0	0.19897
0	0	0.698970004	0	0.1747425
0	0	0.795880017	0	0.19897
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425

0	0	0.698970004	0	0.1747425
0	0	0.698970004	0	0.1747425
0	0	0.795880017	0	0.19897
0	0	0.795880017	0	0.19897
0	0	0	0.698970004	0.1747425
0	0	0	0.698970004	0.1747425
0	0	0	0.665546249	0.16638656
0	0	0	0.665546249	0.16638656
0	0	0	0.795880017	0.19897
0	0	0	0.698970004	0.1747425
0	0	0	0.698970004	0.1747425
0	0	0	0.698970004	0.1747425
0	0	0	0.698970004	0.1747425
0	0	0	0.698970004	0.1747425
0	0	0	0.698970004	0.1747425
0	0	0	0.698970004	0.1747425
0	0	0	0.698970004	0.1747425
0	0	0	0.698970004	0.1747425
0	0	0	0.698970004	0.1747425

Setelah mendapatkan vector posisi untuk centroid 1 dan 2, dihitung jarak perpindahan dari centroid lama ke baru dengan menggunakan rumus euclidian. Setelah dihitung maka di dapat hasil sebagai berikut:

Tabel III.15 Jarak Centroid Lama dan Centroid Baru

Centroid		Jarak
centroid 1 Lama	Centroid 1 baru	4.579092
centroid 2 Lama	Centroid 2 Baru	0.664702

Setelah melakukan perhitungan jarak centroid baru dan lama, Kemudian dilihat apakah jarak antara centroid lama dengan centroid baru lebih dari threshold yang sudah ditentukan di awal.

Jarak C1LamaC1Baru > threshold

4.579092 > 20/100

4.579092 > 0.5

Karena bernilai benar, maka iterasi tidak dilanjutkan dan hasil akhir cluster adalah:

Tabel III.17. Hasil Clustering

Dokument	Jarak	Cluster
D(1)	**4.18938**	2
D(2)	**4.50913**	2
D(3)	**0**	1
D(4)	**5.28548**	2
D(5)	**0**	2

III.1.4.4. Analisis Outlier

Deteksi oulier adalah suatu teknik untuk mencari obyek dimana obyek tersebut mempunyai perilaku yang berbeda dibandingkan obyek-obyek pada umumnya. Teknik data mining dapat digunakan untuk mendeteksi adanya suatu outlier pada sebuah dataset. Teknik data mining yang diganakan adalah Clustering-based, Distance-based dan Density-based.

Dalam tugas akhir ini, untuk mencari dokumen – dokumen yang terdeteksi sebagai outlier menggunakan algoritma *K-Nearest Neighbor*. Hasil clustering dari tahapan sebelumnya akan dicari dokumen – dokumen yang mempunyai outlier.

1. D1 akan dihitung jarak terhadap semua dokumen yang sudah dicluster

Tabel III.18. Hasil clustering

Dokumen	sistem	otomotif	algoritma	Cluster
D1	0.18	0.05	0.03	2
D2	0.01	0	0.08	2
D3	0.07	0.05	0.04	1
D4	0.04	0.16	0.07	2
D5	0.11	0	0	2

Tabel III.19. Dokumen Uji

Dbaru	0.18	0.05	0.03

Pada analisis outlier menggunakan algoritma KNN, dokumen uji adalah dokumen yang ada di dalam dokumen yang sudah dicluster dan akan di uji kembali apakah benar dokumen yang diuji tersebut tepat di cluster tersebut melalui jarak yang dengan menggunakan rumus *euclidien distance*.

Tabel III.20. Jarak dokumen dengan dokumen uji

Dokumen	Jarak = d(Dbaru,D)	Kluster
D1	0.00	2
D2	0.18	2
D3	0.11	1
D4	0.18	2
D5	0.09	2

Setelah dihitung jarak dokumen uji dengan semua dokumen selanjutnya di urutkan nilai yang paling kecil sampai terbesar.

Tabel III.21. Pengurutan Jarak

Dokumen	pengurutan jarak	Kluster
D1	0	2
D5	0.09	2
D3	0.11	1
D2	0.18	2
D4	0.19	2

Dari table diatas di ambil 3 dokumen dengan jarak terkecil untuk dibandingkan, maka di dapat

Tabel III.22. Dokumen jarak terkecil

Dokumen	pengurutan jarak	Kluster
D1	0	2
D5	0.09	2
D3	0.11	1

Table diatas menghasilkan 2 dokumen berada di cluster 2 dan 1 dokumen berada di cluster 1 karena lebih banyak hasil cluster 2 maka, dokumen uji masuk ke cluster 2. Maka dalam hal ini, dokumen uji tersebut adalah D1 **tidak tergolong outlier** karena tidak berpindah cluster karena disebut oulier ketika dokumen tersebut berpindah cluster.

Karena D1 tidak tergolong outlier, maka di lanjutkan ke dokumen selanjutnya dan diuji ke semua dokumen seperti langkah – langkah diatas.

2. D3 akan dihitung jarak terhadap semua dokumen yang sudah dicluster

Tabel III.23. Dokumen Uji

Dbaru	0.07	0.05	0.04

Pada analisis outlier menggunakan algoritma KNN, dokumen uji adalah dokumen yang ada di dalam dokumen yang sudah dicluster dan akan di uji kembali apakah benar dokumen yang diuji tersebut tepat di cluster tersebut melalui jarak yang dengan menggunakan rumus *euclidien distance*.

Tabel III.24. Pengurutan Jarak

Dokumen	Jarak = d(Dbaru,D)	Cluster
D3	0	1
D5	0.08	2
D2	0.09	2
D1	0.11	2
D4	0.12	2

Dari table diatas di ambil 3 dokumen dengan jarak terkecil untuk dibandingkan, maka di dapat

Tabel III.25. Dokumen jarak terkecil

Dokumen	Jarak = d(Dbaru,D)	Cluster
D3	0	1
D5	0.08	2
D2	0.09	2

Table diatas menghasilkan 2 dokumen berada di cluster 2 dan 1 dokumen berada di cluster 1 karena lebih banyak hasil cluster 2 maka, dokumen uji masuk ke cluster 2. Maka dalam hal ini, dokumen uji tersebut adalah D3 disebut **outlier** karena berpindah cluster.

Dari percobaan diatas terdapat dokumen yang outlier, lebih rinci ditunjukkan oleh table III.27 dibawah ini

Tabel III.26. Hasil Analisis Outlier

Dokumen	Cluster	Outlier
D1	2	2
D2	2	2
D3	1	2
D4	2	2
D5	2	1

III.1.5. Analisis Kebutuhan Non-Fungsional

Analisis kebutuhan non fungsional dilakukan untuk mengetahui spesifikasi kebutuhan non fungsional untuk sistem. Spesifikasi kebutuhan non fungsional melibatkan analisis kebutuhan perangkat keras (*hardware*), analisis kebutuhan perangkat lunak (*software*) dan analisis kebutuhan pengguna (*user*).

III.1.5.1. Analisis kebutuhan perangkat keras

Berdasarkan hasil pengamatan, spesifikasi perangkat keras yang saat ini digunakan oleh Pusat Penelitian Informatika – LIPI secara umum adalah sebagai berikut :

1) Processor : Intel Core 2 Duo 2,10 GHz
2) Harddisk : 80 Gb
3) Memory : 1 GB
4) Monitor : 17” LCD
5) Keyboard
6) Optical Mouse
7) DVD RW

Sedangkan kebutuhan perangkat keras (hardware) yang dibutuhkan oleh perangkat lunak yang akan dibangun minimal memiliki spesifikasi sebagai berikut:

1) Processor : Intel Pentium 4 @ 2,20 GHz
2) Harddisk : 40 Gb
3) Memory : 1 GB
4) Monitor
5) Keyboard
6) Optical Mouse

Melihat spesifikasi perangkat keras yang dimiliki Pusat Penelitian Informatika – LIPI sudah memenuhi spesifikasi minimum perangkat keras yang dibutuhkan oleh perangkat lunak pengelompokan artikel ilmiah yang akan dibuat.

III.1.5.2. Analisis kebutuhan perangkat lunak

Berdasarkan hasil pengamatan, perangkat lunak (software) yang saat ini digunakan oleh Pusat Penelitian Informatika – LIPI secara umum adalah sebagai berikut :

1) Sistem Operasi : Windows 7 ultimate 32bit
2) Software Pendukung : Microsoft office 2010

Sedangkan perangkat lunak (software) yang saat ini digunakan dalam membuat aplikasi pengelompokan artikel ilmiah adalah sebagai berikut:

1) Sistem Operasi : Windows 8.1 Pro 32bit
2) Java
3) MySQL 32bit

Spesifikasi perangkat lunak yang dibutuhkan untuk menjalankan aplikasi pengelompokan artikel ilmiah yang akan dibangun adalah sebagai berikut:

1) Sistem Operasi : Windows XP Professional 32bit
2) MySQL 32bit

Dengan membandingkan spesifikasi Perangkat lunak yang digunakan Pusat Penelitian Informatika – LIPI dengan spesifikasi perangkat lunak yang dibutuhkan untuk menjalankan aplikasi yang akan dibangun masih belum memadai, tetapi permasalahan ini bisa diatasi dengan menginstal MySQL 32bit pada komputer yang hendak digunakan.

III.1.5.3. Analisis kebutuhan pengguna

Analisis kebutuhan pengguna dimaksudkan untuk mengetahui karakteristik pengguna yang dapat menggunakan aplikasi yang akan dibangun.

Adapun untuk mengetahui karakteristik pengguna, akan dijelaskan pada table berikut:

Tabel III.27. Analisis Kebutuhan Pengguna Pada Sistem

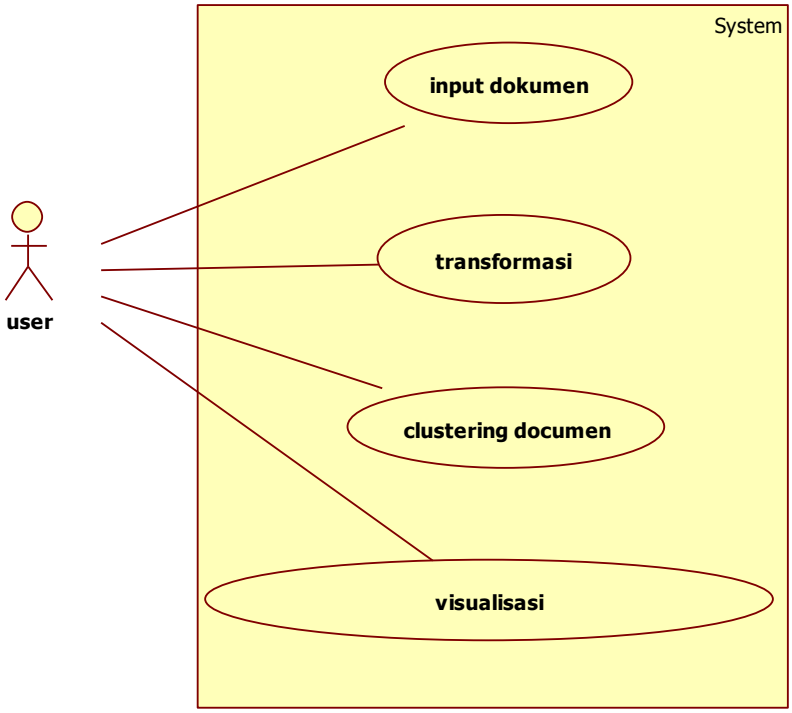
Penggguna	**Tanggung Jawab**	**Hak Akses**	**Tingkat Pendidikan**	**Tingkat Keterampilan**	**Pengalaman**
Kepala Pusat penelitian informatika – LIPI	Mengecek kefalidan artikel ilmiah dan melakukan evaluasi untuk pengembanga	Mengakses artikel ilmiah dan melakukan pengelom	Minimal S1	Mengerti perintah – perintah pada computer dan dapat mengoperasikannya	Pernal melakukan analisis data dan membuat laporan

	n penelitian yang ada di Pusat Penelitian Informatika – LIPI	pkan artikel ilmiah			

III.1.6. Analisis Kebutuhan Fungsional

Analisis kebutuhan fungsional menggambarkan proses kegiatan yang akan diterapkan ke dalam sebuah pengelompokan artikel ilmiah yang akan dibangun dan juga menjelaskan agar sistem dapat berjalan dengan baik serta sesuai dengan kebutuhan. Pemodelan sistem menggunakan pemodelan berorientasi objek dan dimodelkan menggunakan UML (*Unified Modeling Language*). Diagram-diagram UML tersebut diantara lain, yaitu use case diagram, scenario use case, activity diagram, sequence diagram dan class diagram.

III.1.6.1. Use case diagram



Gambar.III.1. Use Case Diagram

1. Defenisi Aktor

Definisi actor berfungsi menjelaskan setiap actor yang terdapat pada use case diagram.

User adalah Pengguna atau orang yang melakukan interaksi terhadap sistem sesuai dengan fasilitas yang disediakan oleh sistem.

2. Defenisi use case

Definisi Use Case adalah penjelasan tentang fungsi dari setiap use case yang terdapat pada diagram use case. Berikut adalah definisi dari use case :

1) *Input dokumen* adalah funsionalitas untuk memasukkan dokumen berformat *.pdf ke dalam aplikasi
2) *Transformasi* adalah fungsionalitas untuk mengubah dokumen inputan yang berformat *.pdf ke dalam format *.txt
3) *Clustering dokumen* adalah fungsionalitas untuk mengekstrak kata – kata yang mewakili setiap dokumen dilanjutkan dengan clustering dokumen.
4) *Visialisasi* adalah fungsionalitas untuk menampilkan hasil daripada cluster yang diperoleh

III.1.6.2. Scenario use case

Use case scenario menjelaskan gambaran dari alur sebuah system dimana setiap scenario menggambarkan dari sudut pandang aktor atau piranti yang berinteaksi dengan system secara manual maupun otomatis atau berbagai cara.

1) Use case scenario input document

Use scenario *input document* menggambarkan proses ketika menu yang dilipilih oleh pengguna adalah input document. Berikut adalah proses dari input document

Tabel III.28. Use Case Scenario Input Document

Identifikasi	
No	UC-01
Nama	*Input document*
Tujuan	Aplikasi memasukkan dokumen artikel ilmiah
Deskripsi	Proses input data
Actor	Pengguna
Scenario Utama	
Kondisi Awal	Pengguna berada pada halaman utama
Aksi Actor	Reaksi system
1. Pengguna memilih menu *input dokumen*	

	2.menampilakan *load dir* untuk memilih folder yang berisi artikel ilmiah
Kondisi Akhir	Aplikasi menampilakan semua artikel ilmiah yang dipilih

2) Use case scenario transformasi

Use case scenario *transformasi* menjelaskan setelah dokumen yang sudah di input dilanjutkan untuk transformasi file format *.pdf ke *.txt. berikut adalah proses dari transformasi

Tabel III.29. use case scenario transformasi

Identifikasi	
No	UC-02
Nama	*transformasi*
Tujuan	Aplikasi mengubah format *.pdf ke *.txt
Deskripsi	Proses perubahan format *.pdf ke *.txt
Actor	Pengguna
Scenario Utama	
Kondisi Awal	Pengguna berada pada halaman utama
Aksi Actor	Reaksi system
1. Pengguna memilih menu *transformasi*	
	2. Masuk ke halaman *transformasi*
3. Pengguna memilih *transform*	
Kondisi Akhir	Aplikasi menampilakan semua artikel ilmiah format *.pdf dan hasil transform berformat *.txt

3) Use case scenario clustering document

Use case scenario clustering document menjelaskan tahapan pada analisis clustering yang sudah di rumuskan pada bagian tahap analisis. Berikut adalah proses dari clustering document.

Tabel III.30. use case scenario clustering

Identifikasi	
No	UC-03
Nama	*Clustering document*
Tujuan	Aplikasi mengelompokkan dokumen format *.txt
Deskripsi	Proses pengelompokan dokumen
Actor	Pengguna
Scenario Utama	
Kondisi Awal	Pengguna berada pada halaman utama
Aksi Actor	Reaksi system
1. Pengguna memilih menu *clustering document*	
	2. Masuk ke halaman clustering document
3. Pengguna memilih *cluster doc*	
Kondisi Akhir	Aplikasi document yang dicluster ditampilkan

4) Use case scenario visualisasi

Use case scenario visualisasi menjelaskan hasil akhir dari cluster ditampilkan berupa table – table setiap cluster

Tabel III.31. use case scenario Outlier

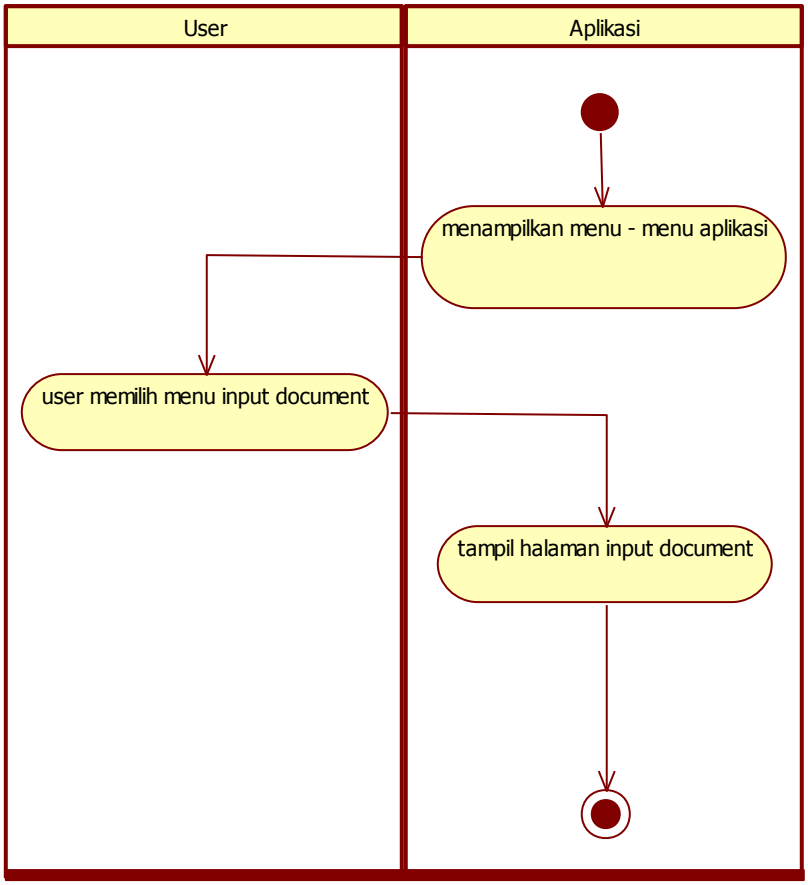
Identifikasi	
No	UC-04
Nama	*visualisasi*
Tujuan	Aplikasi menampilkan hasil cluster
Deskripsi	Proses memvisualisikan hasil cluster
Actor	Pengguna
Scenario Utama	
Kondisi Awal	Pengguna berada pada halaman utama
Aksi Actor	Reaksi system
1. Pengguna memilih menu *visualisasi*	
	2. Masuk ke halaman *visualisasi*
Kondisi Akhir	Aplikasi menampilkan hasil clustering dokumen

III.1.6.3. Activity diagram

Activity diagram menggambarkan berbagai alir aktifitas dalam system yang dirancang, bagaimana masing-masing alir berawal, decision yang mungkin terjadi dan bagaimana mereka berakhir. Penggambaran activity diagram mempunyai kemiripan dengan flowchart diagram. Activity diagram memodelkan event – event yang terjadi pada use case dan digunakan untuk pemodelan aspek dinamis dari sistem.

1) Activity Diagram input document

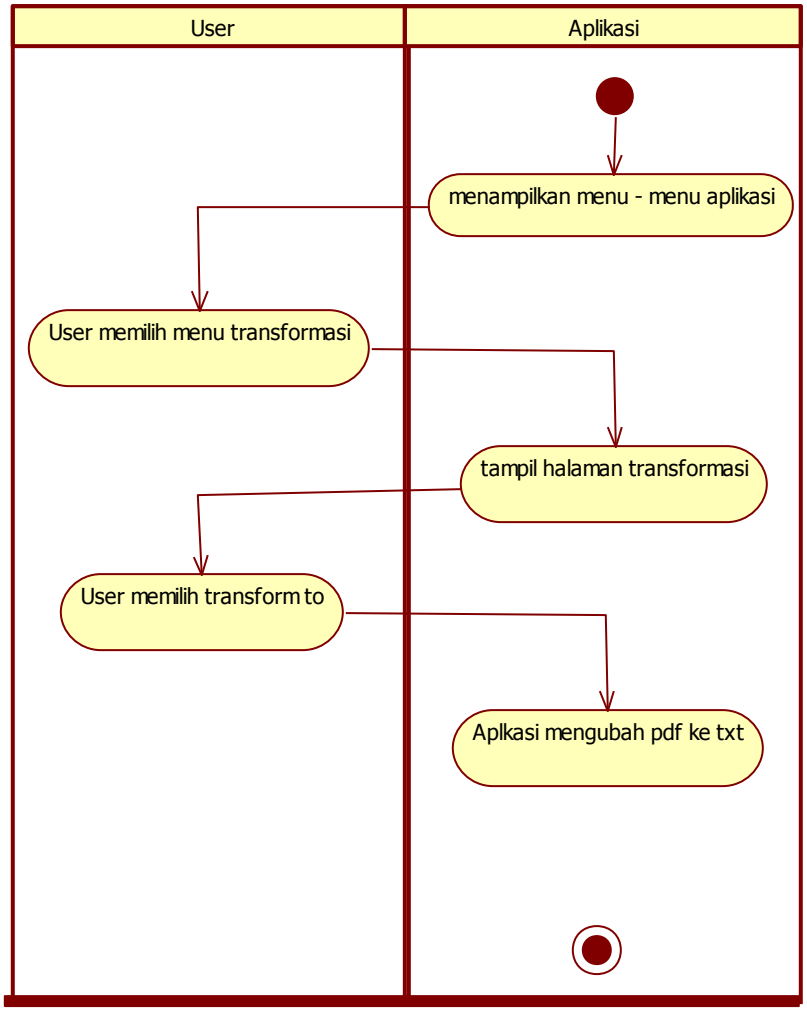
Berikut adalah activity diagram menu input document



Gambar.III.2. activity diagram menu input document

2) Activity Diagram transformasi

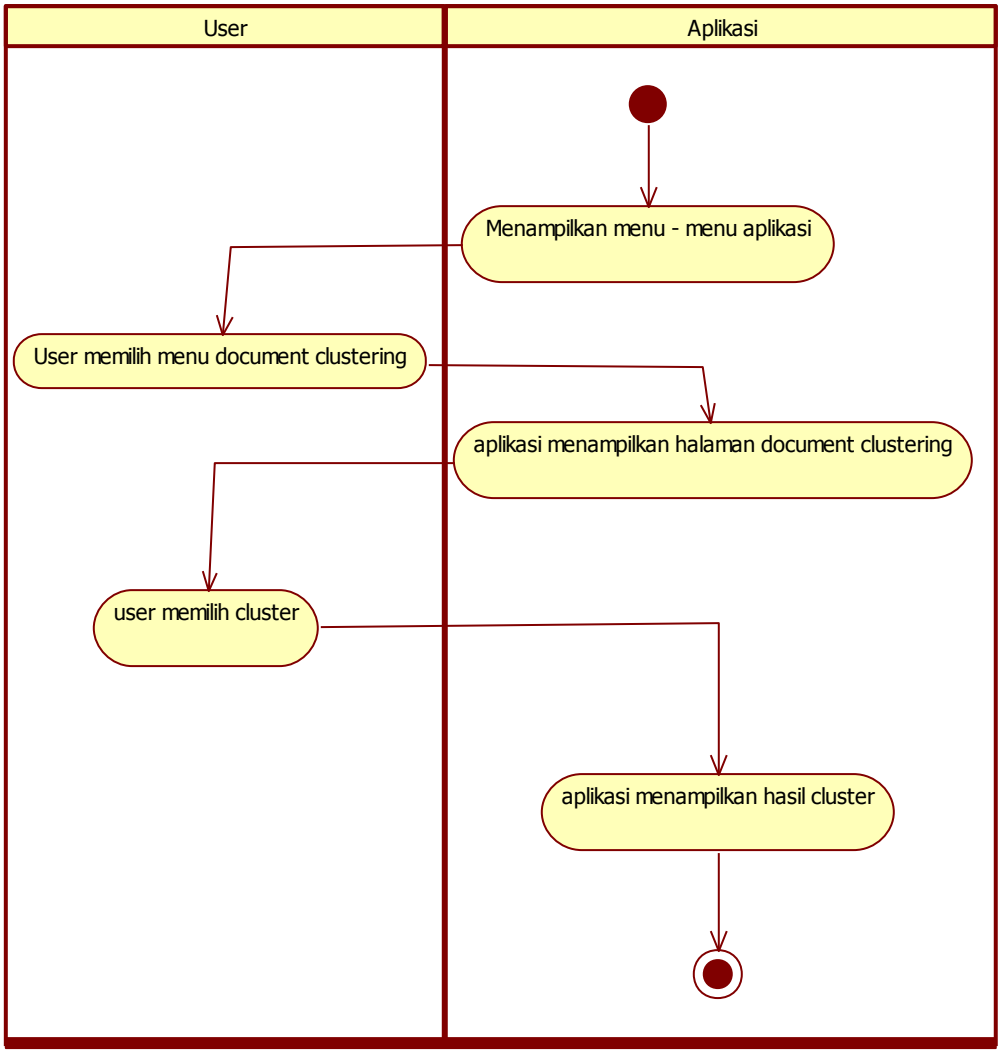
Berikut adalah activity diagram menu transformasi



Gambar.III.3. activity diagram menu transformasi

3) Activity diagram clustering document

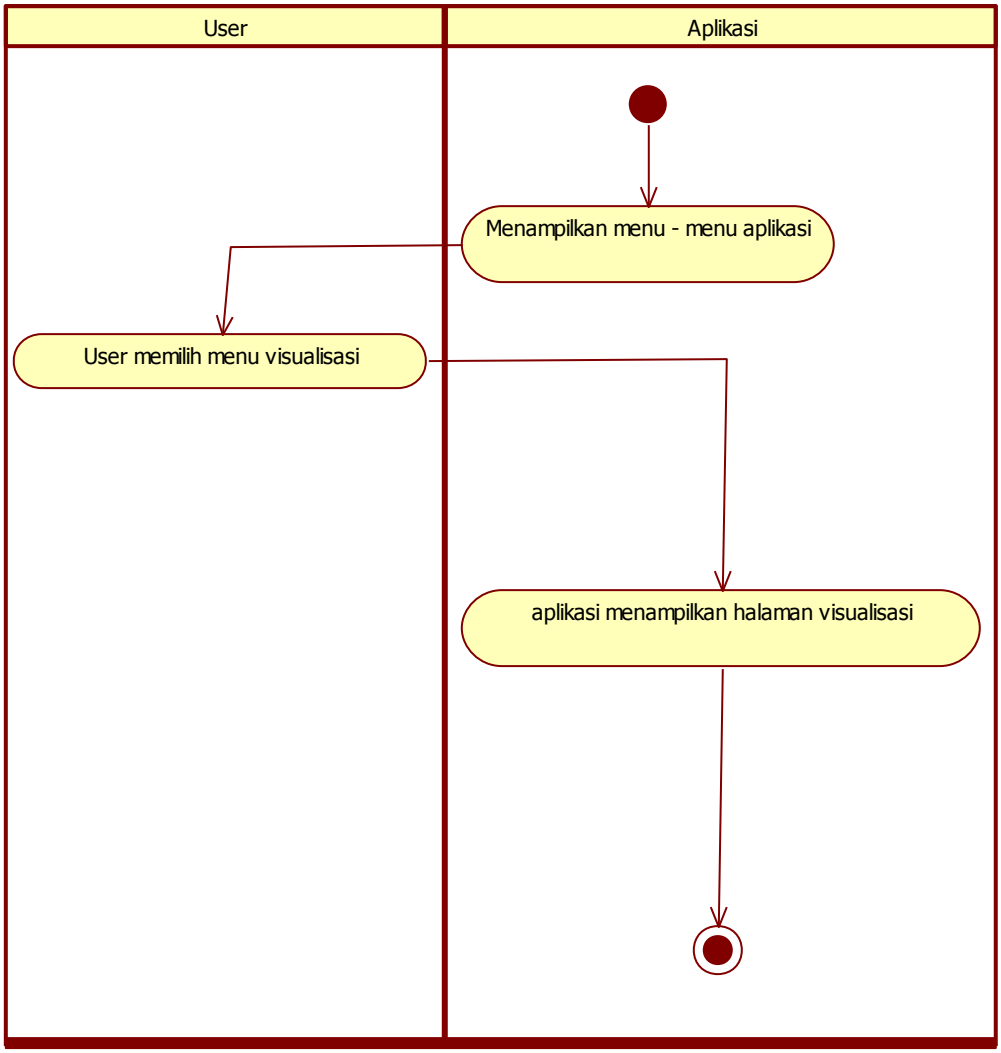
Berikut adalah activity diagram menu clustering document



Gambar.III.4. activity diagram menu Document clustering

4) Activity diagram visualisasi

Berikut adalah activity diagram menu visualisasi



Gambar.III.5. activity diagram menu visualisasi

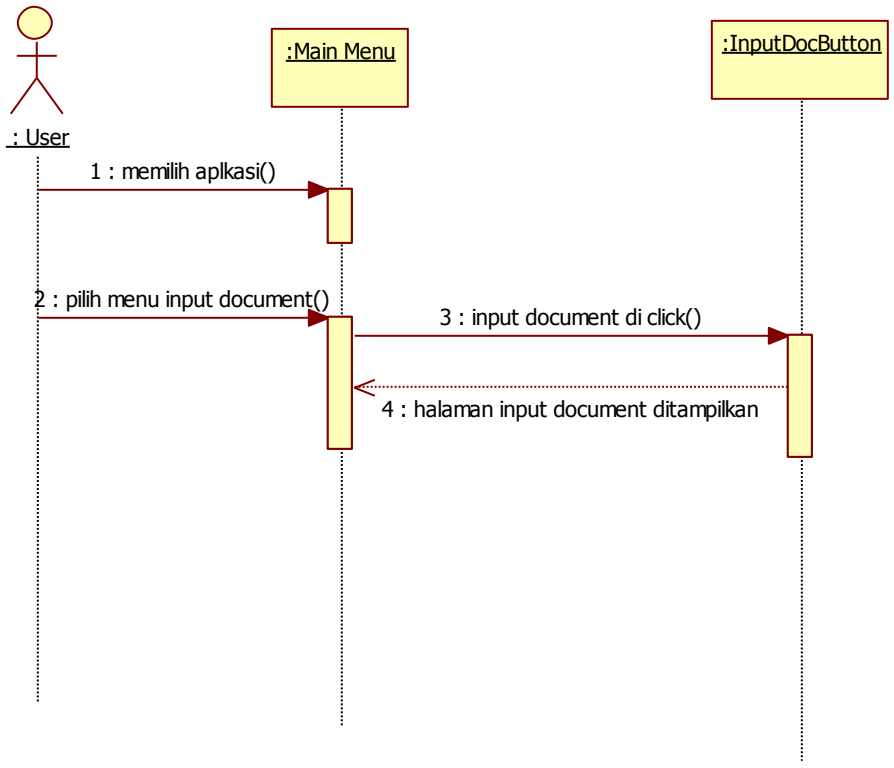
III.1.6.4. Sequence Diagram

Sequence diagram menjelaskan gambaran interaksi antara objek dalam urutan waktu. Interaksi ini berupa pengiriman serangkaian data antara objek-objek yang saling berinteraksi. Dimensi vertikal menjelaskan waktu yang sedang berlangsung sedangkan dimensi horizontal menjelaskan objek yang sedang berhubungan.

Berikut ini merupakan sequence diagram pada aplikasi pengelompokan dokumen artikel ilmiah.

1) Sequence diagram menu input document

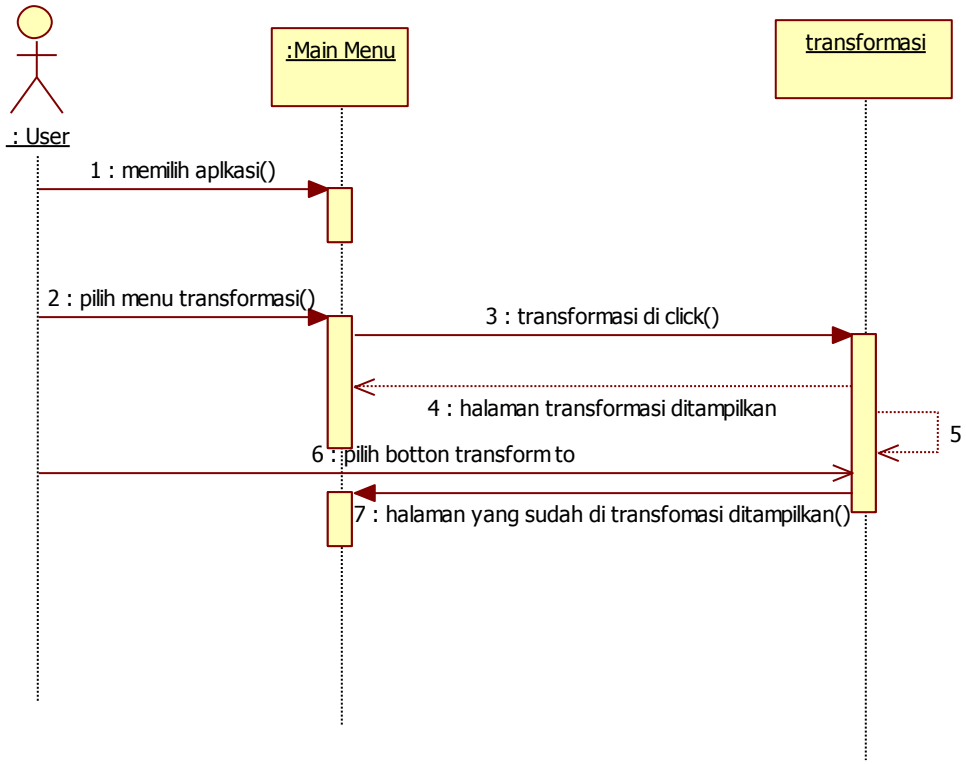
Berikut ini adalah gambaran sequence diagram Menu input document:



Gambar.III.6. sequence diagram menu input document

2) Sequence diagram menu Transformasi

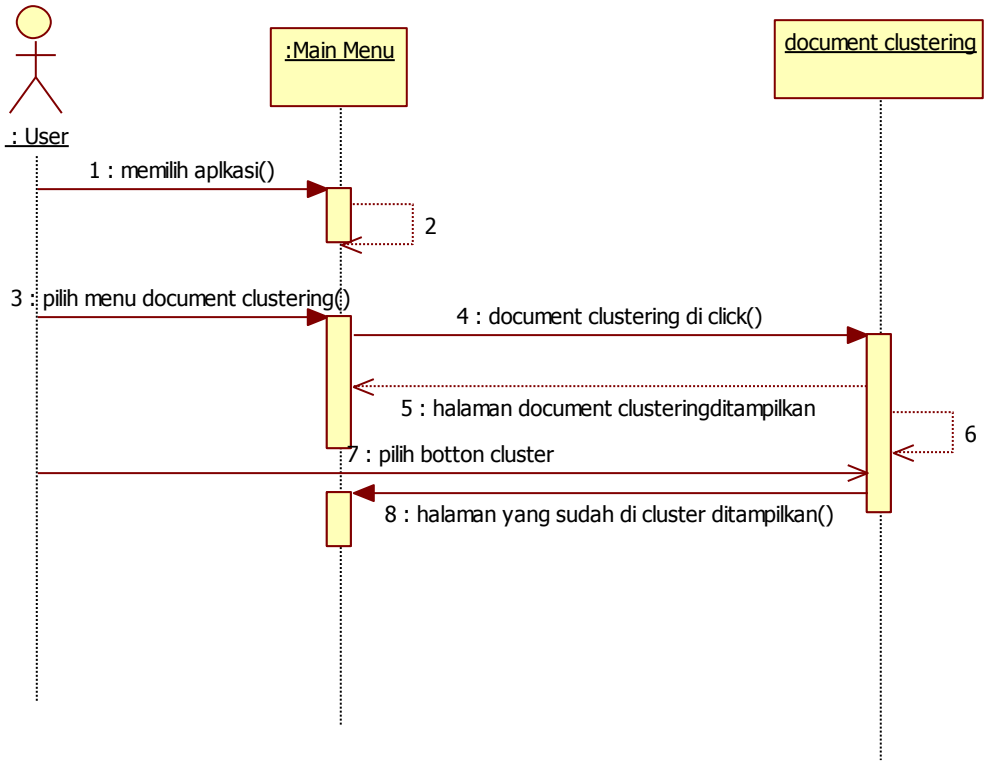
Berikut ini adalah gambaran sequence diagram Menu transformasi:



Gambar.III.7. sequence diagram Menu transformasi

3) Sequence diagram menu document clustering

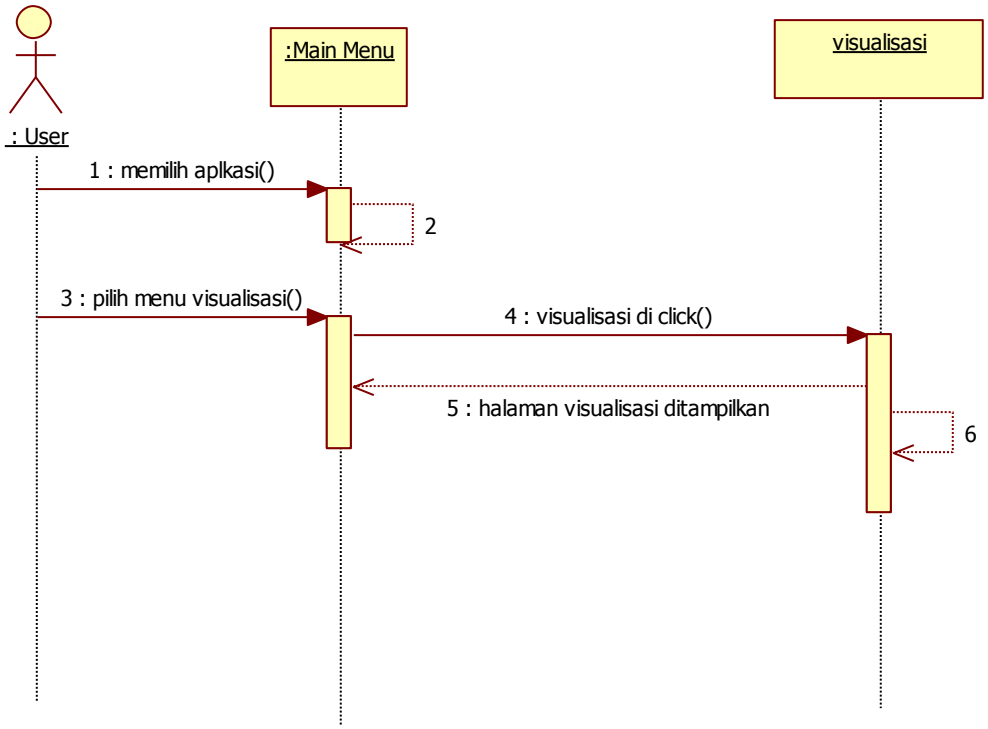
Berikut ini adalah gambaran sequence diagram Menu document clustering:



Gambar.III.8. sequence diagram Menu document clustering

4) Sequence diagram menu visualisasi

Berikut ini adalah gambaran sequence diagram Menu visualisasi:



Gambar.III.9. Sequence Diagram Menu Document Clustering

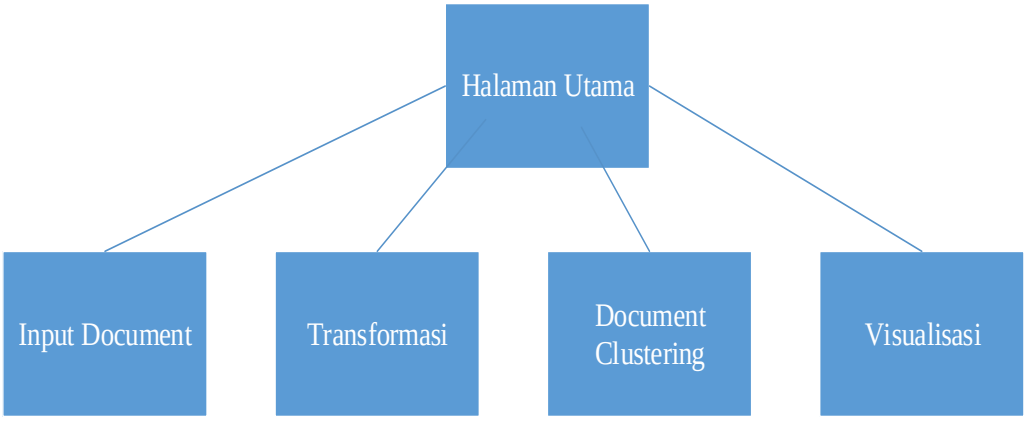
III.2 Perancangan Sistem

Perancangan sistem merupakan suatu bagian dalam sebuah metodologi pengembangan perangkat lunak yang dilakukan setelah melewati tahapan analisis. Perancangan dapat dikatakan sebagai penggambaran, perencanaan dan pembuatan sketsa dari beberapa elemen yang terpisah ke dalam satu kesatuan yang utuh dan berfungsi. Untuk dapat melakukan perancangan sistem maka dulakukan beberapa langkah yaitu sebagai berikut :

1) Perancangan struktur menu
2) Perancangan antarmuka

III.2.1. Perancangan struktur menu

Struktur menu adalah bentuk umum dari suatu rancangan aplikasi untuk memudahkan pengguna dalam menjalankan aplikasi. Berikut ini adalah perancangan struktur menu aplikasi pengelompokan dokumen artikel ilmiah:

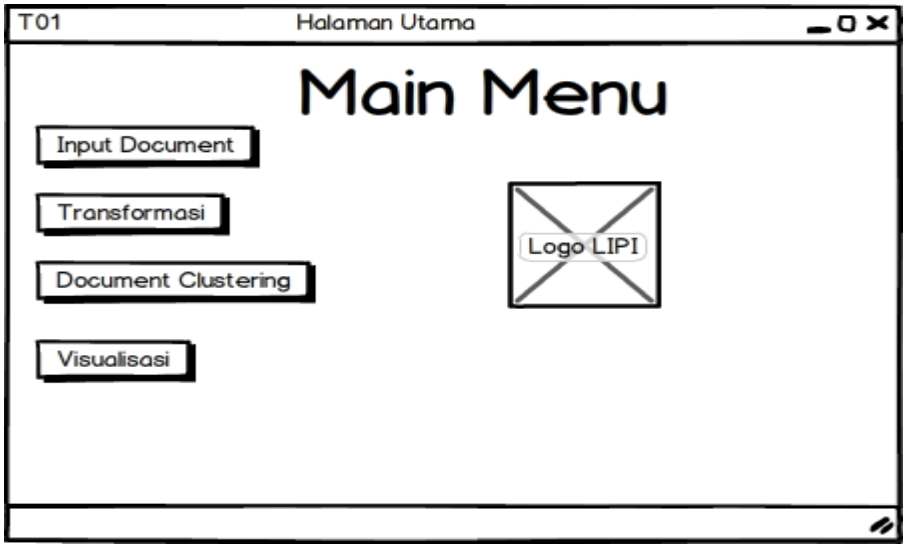


Gambar.III.10. Struktur Menu

III.2.2. Perancangan Antar Muka

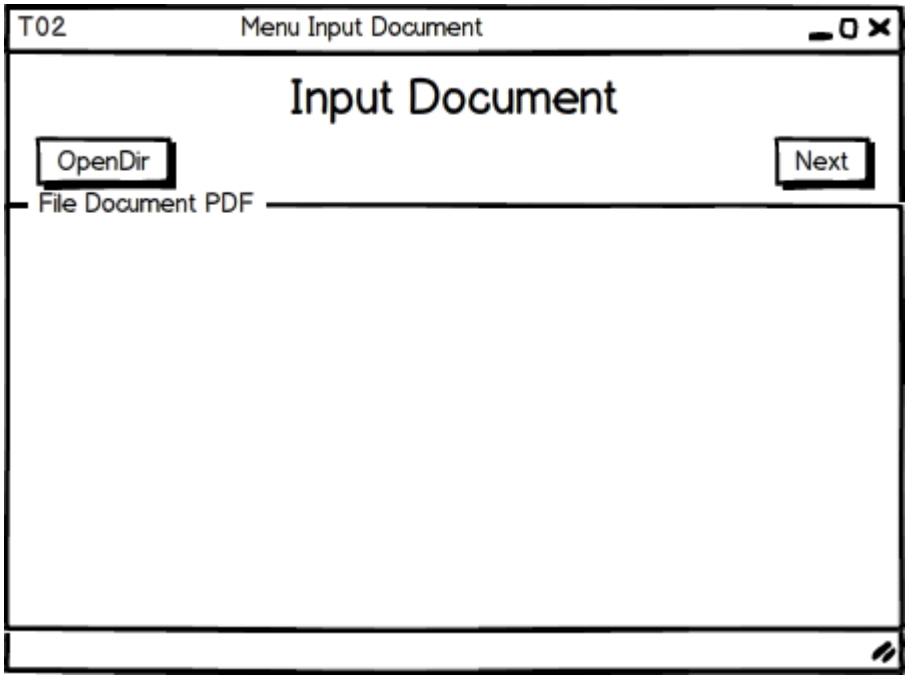
Antarmuka atau interface merupakan tampilan daris suatu program aplikasi yang berperan sebagai media komunikasi yang digunakan antara aplikasi dan pengguna. Antarmuka aplikasi diharapkan mudah dipahami oleh pengguna. Berikut ini adalah gambaran interface pada aplikasi pengelompokan dokumen artikel ilmiah :

1. Perancangan antar muka halaman utama



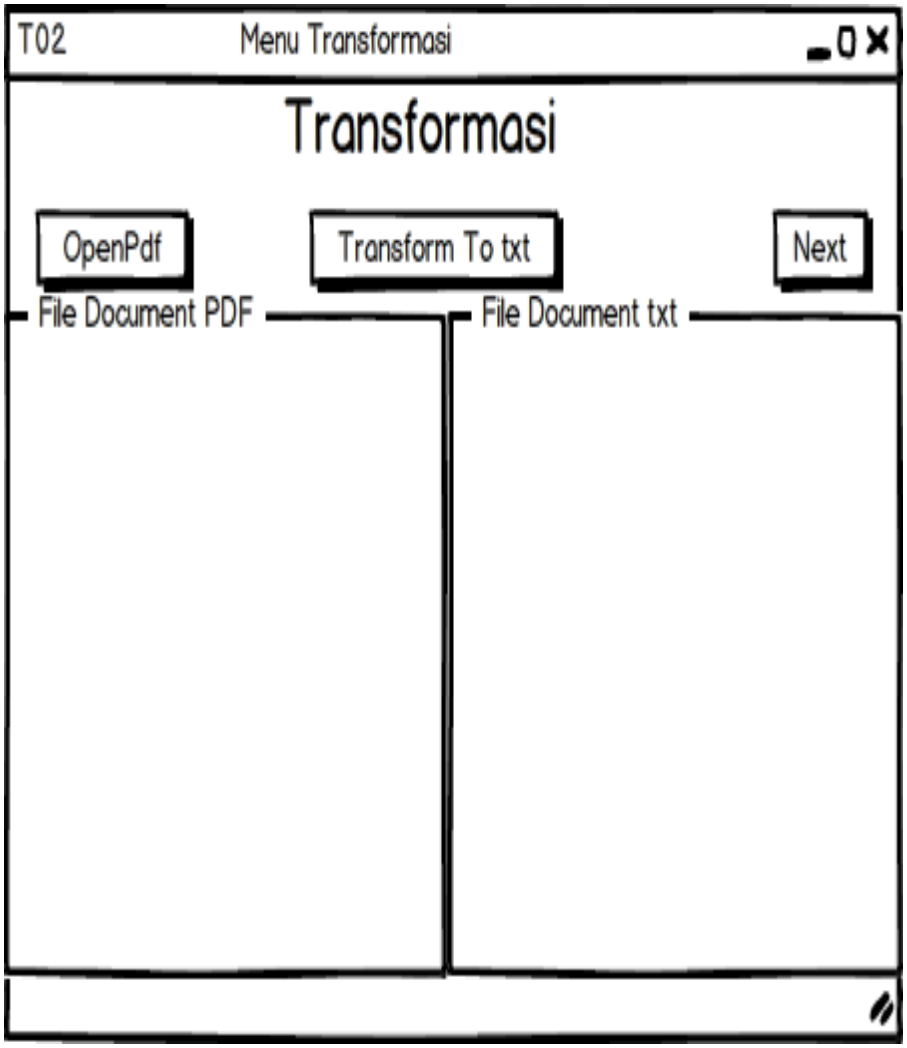
Gambar.III.11. Antarmuka Halaman Utama

2. Perancangan antar muka input document



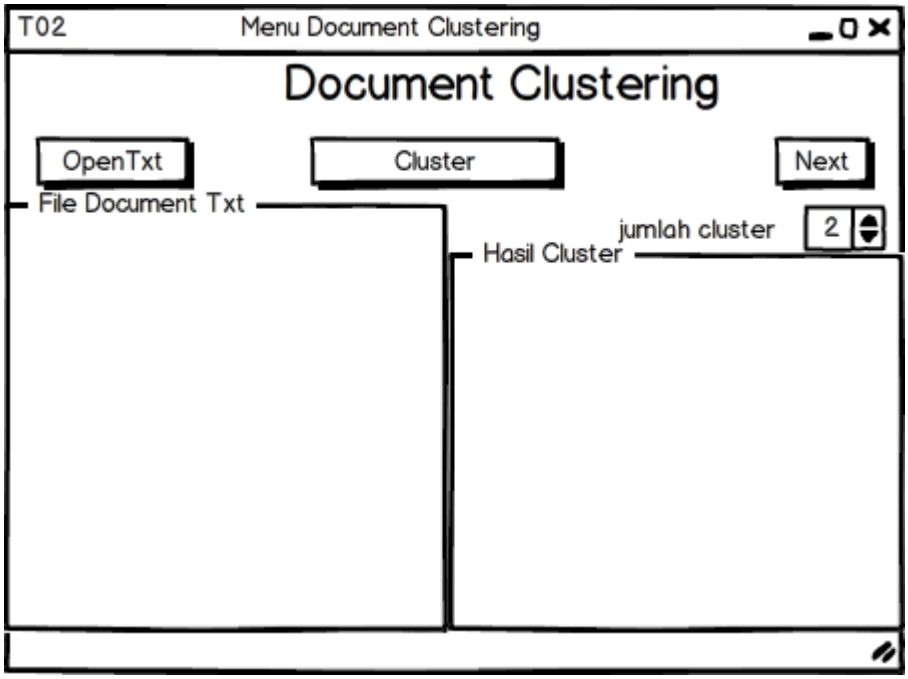
Gambar.III.12. Antarmuka Input Document

3. Perancangan antar muka transformasi



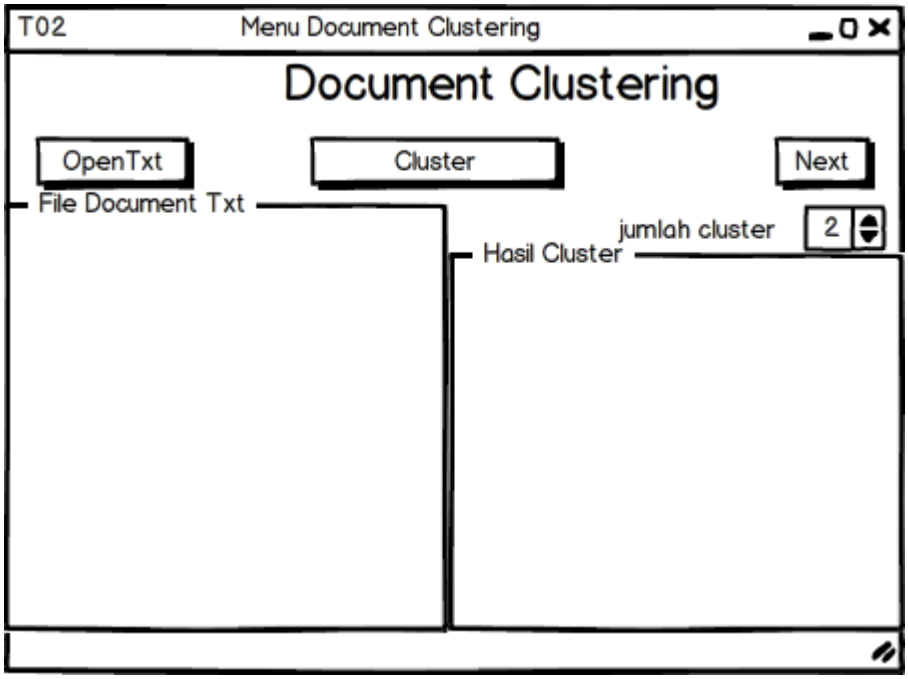
Gambar.III.13. Antarmuka Transformasi

4. Perancangan antar muka document clustering



Gambar.III.14. Antarmuka Transformasi

5. Peranncangan antar muka visualisasi



Gambar.III.15. Antarmuka Transformasi