

BAB III METODOLOGI PENELITIAN

3.1 Metode Pengumpulan Data

Dalam metode pengumpulan data memiliki peranan yang penting dalam mendapatkan suatu informasi. Data yang relevan dengan pokok pembahasannya adalah indikator keberhasilan suatu penelitian. Dalam pengumpulan data, beberapa metode yang digunakan diantaranya:

3.1.1 Wawancara

Wawancara adalah proses pengumpulan data dengan tanya jawab secara langsung dengan orang – orang yang terkait. Berikut ini merupakan beberapa pertanyaan wawancara yang dilakukan yaitu:

- a. Kendala apa saja yang dialami oleh kelompok bidang keilmuan (KBK)?
- b. Berapa lama periode mahasiswa mengumpulkan judul skripsi?

3.1.2 Pengamatan

Penulis mengamati secara langsung lingkungan sistem yang akan dibangun. Ini bertujuan untuk mendapatkan gambaran yang jelas terhadap sistem yang akan dibangun.

3.1.3 Tinjauan Pustaka

Dalam pengumpulan data pada prosesnya dilakukan dengan mempelajari berbagai bentuk bahan tertulis yang berupa buku – buku, artikel, dokumen – dokumen, yang terkait secara langsung dengan sistem kemiripan judul skripsi [15].

3.2 Implementasi Algoritma *TF-IDF* Dan *Fuzzy Matching*

Dalam penelitian ini, untuk melakukan kombinasi algoritma *TF-IDF* dengan algoritma *Fuzzy Matching* terhadap dataset judul-judul skripsi Teknik

Informatika yang berbentuk *spreadsheet* dapat dilakukan dengan langkah-langkah sebagai berikut

3.2.1 Pra-Proses *Dataset*

Langkah pertama yang harus dilakukan ialah pra-proses *dataset*, yang dapat dilakukan dengan cara menghapus data duplikat, menghapus data yang tidak relevan, mengubah huruf-hurufnya menjadi kecil, dan menghapus spasi pada awal dan akhir kalimat [16]. Data redudan dan data tidak valid dapat dilihat pada Tabel 3.1

Tabel 3.1 Data Redudan dan Data tidak Valid

53	03010100	Julian Purnama	APLIKASI ADMINISTRASI SURAT MASUK DAN SURAT KELUAR PADA DINAS PENGAIRAN BANDAR LAMPUNG
54	1711010021	Renaldi	Aplikasi Analisis Penjualan Barang Menggunakan Algoritma Apriori (Studi Kasus : Dayspeed Computer)
55	1311010060	Angga Wahyudi	Aplikasi Android Dengan Algoritma C45 Untuk Menentukan Lama Panen Udang
56	1411010094	Trianto	Aplikasi Audit Sistem Informasi Akademik Institut Informatika dan Bisnis (IIB) Darmajaya Lampung Menggunakan FrameWork COBIT 5
57	1111010133	Rames Krisnan Kuntoro	Aplikasi Automatic Electricity Switch Control Dengan Arduino Berbasis Cloud Pada Smart House System
58	1711010038	Novita Dian Sari	Aplikasi Bahasa Lampung Untuk Siswa Sekolah Dasar Menggunakan Metode ADDIE
59	1411010018	Galuh Wiyenda Muhammad	Aplikasi Bahasa Pemrograman Java Disertai Compiler Berbasis Android (Studi Kasus Pemrograman Menengah IIB Darmajaya)
60	1111010171	Randika Agung Pratama	Aplikasi Bantu Belajar Teknik Beladiri Taekwondo Berbasis Android
61	1211010218	Ori Rahmat Nur	Aplikasi Bantu Belajar Teknik dan Peraturan Dalam Permainan Softball Berbasis Android
62	1211010109	I Made Krishnandra Putra	Aplikasi Bantu Belajar Teknik Dasar Renang Berbasis Android
63		Bitu Parega	Aplikasi Bimbingan Nurul Fikri
64	1911019002P	RIZAL IRIDIAN INDRAS PRASETYA	Aplikasi Chatbot Informasi Kampus STKIP Al Islam Tunas Bangsa dengan menggunakan Metode Artificial Intelegence Markup Language (AIML) Berbasis Android
65	1611010112	Beny Putra A.R	Aplikasi Chatbot Untuk Pelayanan Administrasi Menggunakan Metode Semantic Search Studi Kasus : Kampus lib Darmajaya Bandar Lampung)
66	04010046	Agung Susanto	APLIKASI DAFTAR HADIR SEKRETARIAT KOMISI PEMILIHAN UMUM PROVINSI LAMPUNG
67	1911010025	Rodhi Faisal Mufid	Aplikasi Deteksi Cyber Attack SQL Injection menggunakan machine learning
68	1111010022	Suparmono	Aplikasi Deteksi Jenis Motif Tapis Lampung Menggunakan Metode Sobel
69	1611010151	Evi Kurniasih	Aplikasi Deteksi Manusia dengan Menggunakan Metode Histogram of Oriented Gradient pada BMT Assyafiiyah
70	1411010106	Yuli Astiti	Aplikasi Diagnosa Penyakit Pada Gajah Menggunakan Teorema Bayes Berbasis Android (Studi Kasus Way Kambas Provinsi Lampung)
71	04010212	Risman Simarmata	APLIKASI E-COMMERCE PADA OPTIK MODERN BANDAR LAMPUNG
72	1411010020	Dewi Anggraini	Aplikasi E-Hijab Menggunakan Algoritma Knuth-Morris-Pratt pada Toko Hijab Dailyku Berbasis Android
73	1711010184	Aan Gusnandi	Aplikasi E-Logbook Terhadap Kegiatan Manajemen Perusahaan Berbasis Mobile Pada Pt Mutiara Ferindo Internusa

Pada Tabel 3.1, merupakan dataset yang sudah di hapus data-data yang tidak relevan, dan menghapus data duplikat.

Tabel 3.2 Merubah Huruf-Huruf Menjadi Kecil

index	lower_document
0	pengenalan daun untuk klasifikasi menggunakan image processing
1	rancang bangun aplikasi bela diri muaythay dan winchun berbasis android
2	3d model menstruation cycles sebagai media woman health education
3	account client management dalam jaringan berbasis vlan dan routing di nusatel lampung
4	akses jadwal mata kuliah dengan teknologi wap menggunakan ponsel di darmajaya
5	algoritma dijkstra untuk pencarian rute terpendek pariwisata di

	lampung
6	algoritma genetik untuk optimasi penjadwalan laboratorium
7	algoritma id3 untuk menentukan kelayakan pengajuan kredit mobil di mnc finance bandar lampung
8	analisa algoritma c4.5 untuk memprediksi penjualan pulsa telkomsel pada pt prima serverindo perkasa
9	analisa uji keamanan wpa2 menggunakan fluxion pada pt andaglos global teknologi
dst.	...

Dalam Tabel 3.2, merupakan hasil mengubah huruf-huruf pada dataset menjadi huruf kecil (*lowercase*) sehingga memudahkan proses kalkulasi pada algoritma *TF-IDF* dan *Fuzzy Matching*.

3.2.2 Menghitung Nilai Kemiripan Dengan Algoritma *TF-IDF*

Setelah di proses datasetnya selanjutnya adalah menghitung nilai kemiripannya. Algoritma pertama yang dilakukan dalam perhitungan ini ialah algoritma *TF-IDF* (*Term Frequency-Inverse Document Frequency*). Untuk *query* yang akan digunakan adalah “*perancangan sistem pakar untuk penyakit katarak*”.

Sebelum melakukan perhitungan dengan algoritma *TF-IDF* hal pertama yang harus dilakukan adalah mengubah dokumen ke dalam bentuk vector dengan melakukan tokenisasi.

```

['01' '043' '10' '11' '13' '161' '19' '1912' '1x' '2000' '2016' '21'
'2100sp' '23' '24' '27033' '28' '30' '355' '360' '3d' '4500' '54g' '67'
'98' '_x000d_' 'aan' 'abadi' 'abdul' 'abngun' 'abnormal' 'absensi'
'abses' 'abung' 'ac' 'academic' 'acara' 'accent' 'access' 'account'
'accounting' 'acidindo' 'active' 'acuan' 'acute' 'ad' 'ada' 'adat'
'addie' 'additive' 'address' 'adira' 'admin' 'administrasi'
'administrator' 'adobe' 'aduan' 'advance' 'advent' 'adversarial'
'advertising' 'advokat' 'aero' 'aes' 'afc' 'affine' 'agama' 'agen'
'agent' 'agoritma' 'agri' 'agriculture' 'agrikultur' 'agrindo'
'agromandiri' 'agrotama' 'agung' 'ahp' 'aida' 'aids' 'aiml' 'air' 'aitam'
'aitishop' 'ajar' 'ajaran' 'ajax' 'ajb' 'aji' 'ajoya' 'akademi'
'akademik' 'akar' 'akibat' 'akli' 'akomodasi' 'aksara' 'akses' 'akta'
'aktifitas' 'aktiva' 'aktivitas' 'akuaponik' 'akuntansi' 'akut' 'al'
'alam' 'alang' 'alat' 'album' 'alcatel' 'alergi' 'alfabet' 'alfarina'
'algorita' 'algorithm' 'algoritma' 'algortma' 'algotirma' 'aliran'
'aliun' 'aljabar' 'alkitab' 'alopecia' 'alsha' 'alternatif' 'aluminium'
'alumni' 'alumunium' 'amal' 'amalia' 'ambarawa' 'ambulance' 'amik' 'amil'
'amma' 'ampuh' 'an' 'anak' 'analisa' 'analisi' 'analisis' 'analitical'
'analytc' 'analytic' 'analysis' 'analytic' 'analytical' 'anatomi' 'and']

```

Gambar 3.1 Tokenisasi dokumen

Pada Gambar 1, terlihat proses tokenisasi pada dokumen. Tokenisasi merupakan proses mengubah, memecah urutan teks menjadi unit-unit yang lebih kecil yang disebut token. Token ini dapat berupa kata, frasa, simbol, atau elemen bermakna lainnya [11].

(0, 2880)	0.48259030065799335
(0, 1278)	0.45417268063853794
(0, 2064)	0.15618848224980486
(0, 1665)	0.3771748183224882
(0, 3793)	0.206112276385658
(0, 756)	0.4953685758257398
(0, 2586)	0.3260456960547598
(1, 154)	0.178808224038362
(1, 398)	0.11500576570019669
(1, 3906)	0.4464347265762374
(1, 739)	0.16779056516204108
(1, 2174)	0.4703189810621222
(1, 834)	0.4703189810621222
(1, 378)	0.4703189810621222
(1, 190)	0.16524747299722298
(1, 326)	0.13462826782823697
(1, 2989)	0.13532303702261364
(2, 892)	0.37984729955265784
(2, 1183)	0.4001691271978075
(2, 3914)	0.4001691271978075
(2, 1998)	0.1676347542633631
(2, 3186)	0.1921198627138533
(2, 777)	0.4001691271978075

Gambar 3.2 Proses Vektorisasi

Pada Gambar 3.2, merupakan proses vektorisasi dari proses tokenisasi. Hal ini dilakukan agar kata atau kalimat dapat diproses dalam algoritma *TF-IDF* yang memerlukan vektor angka-angka.

Langkah selanjutnya adalah melakukan *vectorize* terhadap *query* yang dimana *query*-nya adalah “*perancangan sistem pakar untuk penyakit katarak*”.

```
(0, 3793)    0.28711998082389095
(0, 3325)    0.1808093048735593
(0, 2667)    0.2828074406884888
(0, 2636)    0.32330216920724325
(0, 2345)    0.3909757618782555
(0, 1521)    0.7399349460484105
```

Gambar 3.3 Proses Vektorisasi *Query*

Pada Gambar 3.3, dapat terlihat hasil vektorisasi terhadap *query* atau *input*. Selanjutnya setelah didapatkan nilai vektor dari *query* atau *input*, dikalikan dengan vektor *document* yang dibalik atau *Transpose*. *Transpose* merupakan teknik mengubah kolom menjadi baris dan baris menjadi kolom [12].

```
tfidf_scores
abnormal    0.065650
absensi     0.000000
abses       0.000000
abung       0.000000
ac          0.000000
academic    0.000000
acara       0.026461
accent      0.100815
access      0.013534
account     0.012839
accounting  0.000000
acidindo    0.000000
active      0.000000
acuan       0.000000
acute       0.000000
ad          0.000000
ada         0.000000
adat        0.000000
addie       0.000000
additive    0.000000
```

Gambar 3.4 Perkalian Matrix *Query* Dengan *Document*

Pada Gambar 3.4 dapat dilihat hasil dari perkalian *matrix* antara matriks *query* dengan matriks *document* yang menentukan nilai *TF-IDF* nya.

3.2.3 Menghitung Nilai Kemiripan Dengan Algoritma *Fuzzy Matching*

Langkah selanjutnya setelah didapatkan nilai *TF-IDF* nya adalah menghitung nilai *query* terhadap *document* dengan algoritma *Fuzzy Matching* yang nantinya akan dikalikan dengan algoritma *TF-IDF*.

Query: perancangan sistem pakar untuk penyakit katarak

	documents	fuzzy_matching_scores
0	Pengenalan Daun Untuk Klasifikasi Menggunakan...	40
1	Rancang Bangun Aplikasi Bela Diri Muaythay Da...	25
2	3D Model Menstruation Cycles Sebagai Media Wom...	27
3	Account Client Management Dalam Jaringan Berba...	24
4	AKSES JADWAL MATA KULIAH DENGAN TEKNOLOGI WAP ...	29
5	Algoritma Dijkstra Untuk Pencarian Rute Terpen...	36
6	Algoritma Genetik untuk Optimasi Penjadwalan L...	33
7	Algoritma ID3 untuk Menentukan Kelayakan Penga...	33
8	Analisa Algoritma C4.5 Untuk Memprediksi Penju...	40
9	Analisa Uji Keamanan WPA2 Menggunakan Fluxion ...	22
10	ANALISIS ASPEK KEAMANAN SHORT MESSAGE SERVICE ...	21

Gambar 3.5 Nilai Query Terhadap Document Dengan Fuzzy Matching

Pada Gambar 3.5 dapat dilihat hasil perhitungan jarak antara *query* dengan setiap *text* di dalam *document*. Langkah terakhir untuk mengkombinasikan kedua algoritma *TF-IDF* dan *Fuzzy Matching* ialah dengan mengkalikan hasil dari kedua algoritma tersebut.

3.2.4 Kombinasi Algoritma *TF-IDF* dan *Fuzzy Matching*

Setelah didapatkan nilai dari *TF-IDF* dengan *Fuzzy Matching*, langkah selanjutnya adalah mengkombinasikan hasil kedua algoritma tersebut dengan mengkalikan nilainya.

Query: perancangan sistem pakar untuk penyakit katarak

	Judul	TF-IDF x Fuzzy Matching
0	Pengenalan Daun Untuk Klasifikasi Menggunakan...	0.023672
1	Rancang Bangun Aplikasi Bela Diri Muaythay Da...	0.000000
2	3D Model Menstruation Cycles Sebagai Media Wom...	0.000000
3	Account Client Management Dalam Jaringan Berba...	0.000000
4	AKSES JADWAL MATA KULIAH DENGAN TEKNOLOGI WAP ...	0.000000
5	Algoritma Dijkstra Untuk Pencarian Rute Terpen...	0.019322
6	Algoritma Genetik untuk Optimasi Penjadwalan L...	0.019364
7	Algoritma ID3 untuk Menentukan Kelayakan Penga...	0.014086
8	Analisa Algoritma C4.5 Untuk Memprediksi Penju...	0.015770
9	Analisa Uji Keamanan WPA2 Menggunakan Fluxion ...	0.000000
10	ANALISIS ASPEK KEAMANAN SHORT MESSAGE SERVICE ...	0.000000
11	ANALISIS DAN DESAIN JARINGAN PENGONTROL AKTIVI...	0.000000
12	ANALISIS DAN DESAIN WEB E-LEARNING MENGGUNAKAN...	0.000000
13	ANALISIS DAN PERANCANGAN SISTEM INFORMASI AKAD...	0.041598
14	Analisis Dan Perancangan Sistem Informasi Akad...	0.045020

Gambar 3.6 Kombinasi Nilai *TF-IDF* dan *Fuzzy Matching*

Pada Gambar 3.6 terlihat hasil dari kombinasi kedua algoritma, dimana kombinasi dilakukan dengan perkalian sederhana berdasarkan nilai *TF-IDF* dengan *Fuzzy Matching*.

3.3 Analisis Sistem

Dalam perancangan sebuah sistem, diperlukan langkah untuk mengidentifikasi dan mengevaluasi permasalahan atau hambatan dalam proses perancangan. Adapun kebutuhan dalam perancangan yang akan diimplementasikan diantaranya, perangkat keras dan perangkat lunak.

3.2.1 Analisis Kebutuhan Perangkat Lunak

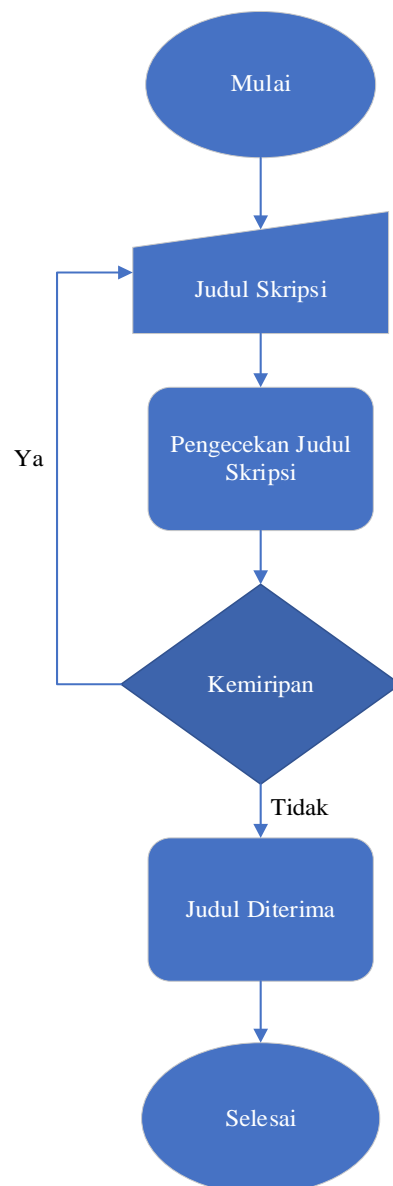
Dalam implementasi aplikasi perangkat lunak, maka software yang dibutuhkan adalah sebagai berikut:

1. Sistem Operasi Windows

2. Bahasa Pemrograman Java
3. Android Studio
4. *Internet*

3.3 Perancangan Sistem

Desain aplikasi dibuat untuk memudahkan dalam pembuatan aplikasi yang berjalan berdasarkan kebutuhan sistem pada alur sistem proses pengecekan kemiripan judul skripsi.



Gambar 3.7 *Flowchart* Proses Pengecekan

Berdasarkan alur pada Gambar 3.7 proses pengecekan kemiripan judul skripsi, dimulai dari memasukkan judul skripsi, selanjutnya sistem akan memproses tingkat kemiripan judul skripsi yang dimasukkan dengan beberapa judul skripsi yang sudah pernah ada. Apabila tingkat kemiripan judul skripsi tinggi, maka akan ditolak terhadap judul yang diajukan oleh mahasiswa, dan apabila tingkat kemiripan tidak terlalu tinggi maka judul memiliki kemungkinan untuk diterima.

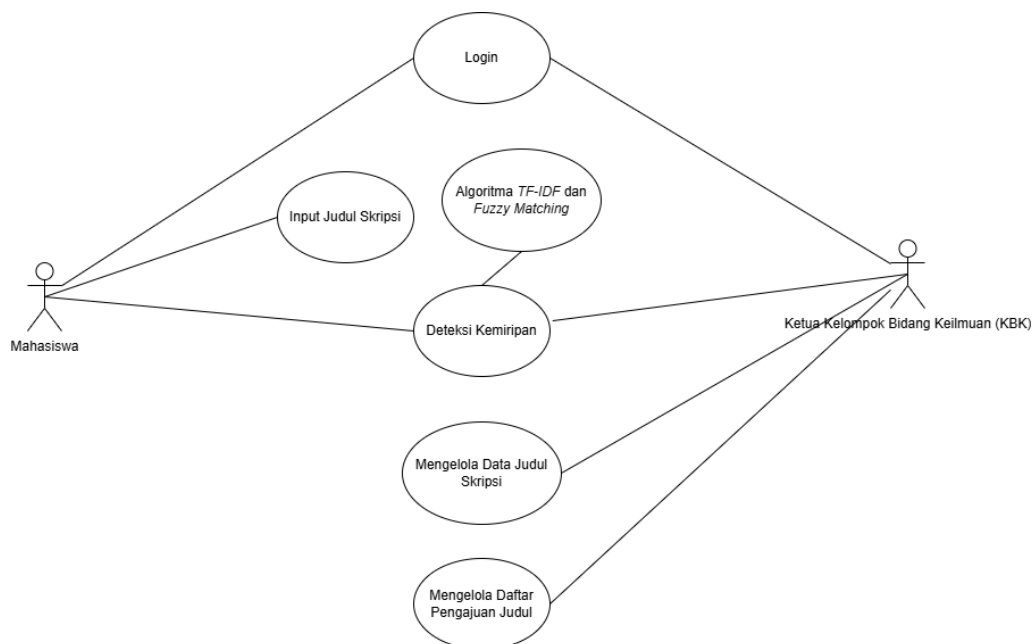
Berdasarkan masalah dan alur sistem yang ada, maka akan dibangun sistem yang dapat mengukur tingkat kemiripan judul skripsi menggunakan algoritma *TF-IDF* dengan kombinasi algoritma *Fuzzy Matching*.

3.4 Tahapan Membangun *Prototype*

Tools yang digunakan dalam perancangan ini adalah UML yaitu penggambaran *use case diagram*, *activity diagram*, dan *class diagram*.

3.4.1 Penggambaran *Usecase Diagram*

Usecase Diagram merupakan *tool* yang mendeskripsikan interaksi antar satu atau lebih aktor dengan sistem informasi yang akan dibuat. Berikut merupakan *usecase diagram* dari sistem deteksi kemiripan judul skripsi pada Gambar 3.8.



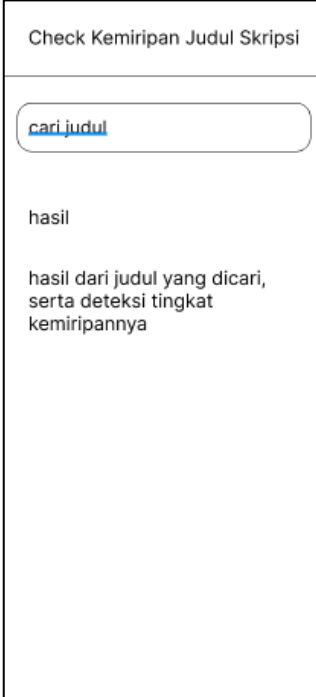
Gambar 3.8 Diagram *Usecase*

3.5 Desain Aplikasi

Desain aplikasi melibatkan pembuatan antarmuka pengguna yang memfasilitasi pengguna dalam mencapai tujuan mereka dengan cara yang efektif dan efisien [13].

3.5.1 Desain Tampilan Awal Aplikasi

Tampilan awal merupakan tampilan yang dihasilkan oleh program saat pertama kali program dijalankan. *prototype* dari tampilan awal aplikasi deteksi kemiripan judul skripsi dapat terlihat pada Gambar 3.9.



The image shows a vertical rectangular window with a black border. At the top, the title "Check Kemiripan Judul Skripsi" is centered. Below the title is a rounded rectangular input field containing the text "cari.judul" in blue. Underneath the input field, the word "hasil" is displayed. Below "hasil", there is a block of text: "hasil dari judul yang dicari, serta deteksi tingkat kemiripannya".

Gambar 3.9 *Prototype* Tampilan Awal Aplikasi

3.5.3 Tampilan Awal Hak Akses Ketua KBK

Check Kemiripan Judul Skripsi

TAMBAH EDIT

cari judul

hasil

hasil dari judul yang dicari,
serta deteksi tingkat
kemiripannya

Gambar 3.10 *Prototype* Tampilan Awal Admin

3.5.4 Tampilan Tambah Judul Hak Akses Ketua KBK

Check Kemiripan Judul Skripsi

Tambah Judul

Nama Judul

Nama Mahasiswa

Tahun Pengajuan

TAMBAH

Gambar 3.11 *Prototype* Tampilan Tambah Judul

3.5.5 Tampilan Edit Judul Hak Akses Ketua KBK

Check Kemiripan Judul Skripsi

Edit Judul

cari judul

Nama Judul

Nama Mahasiswa

Tahun Pengajuan

Gambar 3.12 *Prototype* Tampilan Edit Judul

3.6 Jadwal Penelitian

Tabel 3.3 Jadwal Penelitian

No	Kegiatan	Bulan Ke-1				Bulan Ke-2			
		1	2	3	4	1	2	3	4
1	Wawancara Dan Pengumpulan Data								
2	Merancang dan Mendesain Sistem								
3	Mengembangkan Sistem								
4	Implementasi Sistem								
5	Testing dan Pengujian Sistem								