

**SKRIPSI**

**DETEKSI *PLAGIARISM* PADA  
TUGAS AKHIR MENGGUNAKAN METODE  
*RABIN-KARP* BERBASIS WEB**



Disusun Oleh:

Nama : Ahmad Subadri  
NIM : 17121035

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS MERCU BUANA YOGYAKARTA  
TAHUN 2021**

# **DETEKSI *PLAGIARISM* PADA TUGAS AKHIR MENGGUNAKAN METODE *RABIN-KARP* BERBASIS WEB**

Disusun Oleh:

Nama : Ahmad Subadri  
NIM : 17121035

**Skripsi ini diajukan untuk memenuhi persyaratan akademik  
sarjana pada Program Studi Sistem Informasi, Fakultas  
Teknologi Informasi, Universitas Mercu Buana Yogyakarta.**

Yogyakarta, 13 April 2022  
Menyetujui Pembimbing,

Irfan Pratama S.Kom., M.Eng., MCE.  
NIDN 0529049102

# **HALAMAN PENGESAHAN**

## **DETEKSI *PLAGIARISM* PADA TUGAS AKHIR MENGUNAKAN METODE *RABIN-KARP* BERBASIS WEB**

Oleh:

Ahmad Subadri  
17121035

Telah dipertanggungjawabkan dan  
diterima oleh Tim Penguji pada  
tanggal  
<tgl-bulan-tahun>

Mengetahui Dekan,

Dosen Pembimbing,

(Anief Fauzan Rozi S.Kom.,M.Eng.)  
NIDN. 0522088601

(Irfan Pratama S.Kom., M.Eng., MCE.)  
NIDN. 0529049102

Dosen Penguji,

1. <dosen penguji 1>  
NIDN.

2. <dosen penguji 2>  
NIDN.

3. <dosen penguji 3>  
NIDN.

## **HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS**

Sebagai mahasiswa Universitas Mercu Buana Yogyakarta, saya yang bertanda tangan di bawah ini:

Nama : Ahmad Subadri  
NIM : 17121035  
Program Studi : Sistem Informasi  
Fakultas : Teknologi Informasi  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Mercu Buana Yogyakarta Hak Bebas Royalti Non-eksklusif (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

### **DETEKSI *PLAGIARISM* PADA TUGAS AKHIR MENGGUNAKAN METODE *RABIN-KARP* BERBASIS WEB**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Royalti Non-eksklusif ini Universitas Mercu Buana Yogyakarta berhak menyimpan, mengalih-media-kan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : YOGYAKARTA  
Pada Tanggal : <tgl-bulan-tahun>

Yang menyatakan,

(Ahmad Subadri)

## HALAMAN PERNYATAAN ORISINALITAS KARYA

**Skripsi ini adalah hasil karya saya sendiri,  
dan semua sumber baik yang dikutip maupun yang  
dirujuk, telah saya nyatakan dengan benar.**

Nama Lengkap : Ahmad Subadri  
NIM : 17121035  
Judul Skripsi : Deteksi *plagiarism* pada tugas akhir  
menggunakan metode *Rabin-Karp* berbasis  
web  
Program Studi : Sistem Informasi  
Tanggal : <tgl-bulan-tahun>  
Tanda Tangan : <ttd>

## **HALAMAN PERSEMBAHAN**

Tuliskan beberapa kalimat puitis persembahkan karya tulis ini untuk beberapa orang terdekat dengan Anda.

Format tulisan bebas, termasuk jenis, ukuran.

## **HALAMAN MOTTO**

Tuliskan motto Anda.

Format tulisan bebas, termasuk jenis, ukuran.

## KATA PENGANTAR

Paragraf pertama berisi ungkapan rasa syukur atas terselesainya karya tulis ini.

Selanjutnya dapat dituliskan ucapan terima kasih kepada beberapa pihak yang terkait. Diawali dengan kalimat pembuka, kemudian tuliskan beberapa pihak yang tersebut dalam pointer angka arab (1,2,3,4,...).

Contoh:

1. Pihak pertama, sebagai....., atas bantuan dan kerjasamanya dalam hal.....
2. Pihak pertama, sebagai....., atas bantuan dan kerjasamanya dalam hal.....
3. Pihak pertama, sebagai....., atas bantuan dan kerjasamanya dalam hal.....
4. Pihak pertama, sebagai....., atas bantuan dan kerjasamanya dalam hal.....
5. Pihak pertama, sebagai....., atas bantuan dan kerjasamanya dalam hal.....

Paragraf terakhir adalah paragraf penutup. Beri ungkapan seperlunya.

Yogyakarta, 13 April 2022  
Penulis,

Ahmad Subadri



## **Deteksi *plagiarism* pada tugas akhir menggunakan metode *Rabin-karp* berbasis web**

1. Peneliti 1 (Nama Mhs, NIM, email)
2. Peneliti 2 (Nama Dosen Pembimbing, NIDN, email)

### **ABSTRAK**

Abstrak berisikan ringkasan yang terdiri dari 4 paragraf, yaitu fokus penelitian, metodologi penelitian, hasil penelitian yang di analisis dan dibahas, dan kesimpulan. Lihat kembali aturan penulisan abstrak dalam panduan skripsi mengenai ukuran tulisan dan jarak antar baris.

Paragraf pertama yaitu fokus penelitian. Paragraf ini berisi problem yang dihadapi dengan sistem yang ada pada saat penelitian dilakukan.

Paragraf kedua yaitu metodologi penelitian. Bagian ini memuat uraian tentang langkah-langkah penyelesaian masalah selama melaksanakan penelitian. Langkah- langkah tersebut harus disesuaikan dengan judul penelitian. Misalnya dengan eksperimen, kuesioner, studi kasus.

Paragraf ketiga yaitu analisis dan pembahasan. Paragraf ini memuat hasil penelitian yang kemudian dianalisis dan dibahas.

Paragraf keempat yaitu kesimpulan. Bagian ini berisi garis besar penelitian dan usulan penelitian yang akan datang.

Kata kunci: fokus penelitian, metodologi penelitian, analisis dan pembahasan, kesimpulan.

## DAFTAR ISTILAH

Tuliskan dalam bentuk tabel dengan formaturut abjad menaik (A-Z).

No.	Istilah	Keterangan

# DAFTAR ISI

<b>SKRIPSI .....</b>	<b>i</b>
<b>DETEKSI <i>PLAGIARISM</i> PADA TUGAS AKHIR MENGGUNAKAN METODE <i>RABIN-KARP</i> BERBASIS WEB.....</b>	<b>i</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>ii</b>
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....</b>	<b>iii</b>
<b>HALAMAN PERNYATAAN ORISINALITAS KARYA.....</b>	<b>iv</b>
<b>HALAMAN PERSEMBAHAN.....</b>	<b>v</b>
<b>HALAMAN MOTTO.....</b>	<b>vi</b>
<b>KATA PENGANTAR .....</b>	<b>vii</b>
<b>Deteksi <i>plagiarism</i> pada tugas akhir menggunakan metode <i>Rabin-karp</i> berbasis web.....</b>	<b>viii</b>
<b>ABSTRAK .....</b>	<b>viii</b>
<b>DAFTAR ISTILAH.....</b>	<b>ix</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL.....</b>	<b>xii</b>
<b>DAFTAR GAMBAR .....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
<b>1.1 Latar Belakang.....</b>	<b>1</b>
<b>1.2 Perumusan Masalah.....</b>	<b>3</b>
<b>1.3 Tujuan Penelitian.....</b>	<b>3</b>
<b>1.4 Manfaat Penelitian.....</b>	<b>2</b>
<b>BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....</b>	<b>2</b>
<b>2.1 Tinjauan Pustaka.....</b>	<b>2</b>
<b>2.2 Landasan Teori .....</b>	<b>2</b>
2.2.1 Plagiarisme .....	2
2.2.2 Algoritma Rabin Karp.....	2
2.2.3 Preprocessing.....	2
2.2.4 Case Folding.....	2
2.2.5 Filtering.....	2
2.2.6 Synonym Recognition.....	2
2.2.7 Hashing .....	14
2.2.8 Parsing K-grams .....	14
<b>BAB III METODOLOGI.....</b>	<b>14</b>
<b>3.1 Bahan Penelitian .....</b>	<b>14</b>

<b>3.2</b>	<b>Alat Penelitian .....</b>	<b>14</b>
<b>3.3</b>	<b>Jalan Penelitian .....</b>	<b>14</b>
3.3.1	Metode Pengembangan Perangkat Lunak .....	14
3.3.2	Perancangan Proses .....	14
3.3.3	Skenario Eksperimen .....	14
<b>BAB IV</b>	<b>ANALISIS DAN PEMBAHASAN .....</b>	<b>17</b>
<b>4.1</b>	<b>Hasil Penelitian.....</b>	<b>17</b>
<b>4.2</b>	<b>Analisis dan Pembahasan.....</b>	<b>17</b>
<b>BAB V</b>	<b>PENUTUP .....</b>	<b>18</b>
<b>5.1</b>	<b>Kesimpulan.....</b>	<b>18</b>
<b>5.2</b>	<b>Saran.....</b>	<b>18</b>
<b>DAFTAR PUSTAKA</b>	<b>.....</b>	<b>19</b>
<b>LAMPIRAN.....</b>	<b>.....</b>	<b>20</b>
	Lampiran A Biodata Peneliti.....	20
	Lampiran C Artifak/Data Penelitian.....	22
	Lampiran D Kode Program.....	23

## **DAFTAR TABEL**

## DAFTAR GAMBAR

Gambar 3. 1 Siklus Hidup Metode Pengembangan eXtreme Programming (Ambler, 2002) .....	7
Gambar C. 1 <i>Scan</i> Formulir Pendaftaran .....	xiv
Gambar C. 2 <i>Scan</i> Surat Pernyataan Kesanggupan Pembayara Biaya Administrasi .....	xiv
Gambar C. 3 Diagram Alir Prosedur Pendaftaran .....	xv
Gambar D. 1 Kode Program Form Pendaftaran .....	xvi
Gambar D. 2 Kode Program Proses Penyimpanan Data ke <i>Database</i> .....	xvi

## DAFTAR LAMPIRAN

Lampiran A Biodata Peneliti.....	xii
Lampiran B <i>Scan</i> Dokumen Bimbingan Skripsi .....	xiii
Lampiran C Artifak/Data Penelitian.....	xiv
Lampiran D Kode Program.....	xvi

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang.**

Perkembangan teknologi saat ini sangat pesat dan perannya sangat penting untuk mendukung aktivitas manusia agar dapat mengoptimalkan waktu dengan lebih baik. Salah satu contoh teknologi yang membantu aktivitas manusia adalah sistem informasi. Instansi yang bergerak di bidang pendidikan seperti universitas juga membutuhkan dukungan sistem informasi dalam peningkatan mutu pelayanan terhadap proses pendidikan.

Tugas akhir atau biasa disebut dengan skripsi merupakan istilah yang digunakan untuk menggambarkan suatu karya tulis ilmiah berupa tulisan hasil penelitian seorang sarjana yang disusun untuk memenuhi persyaratan memperoleh gelar sesuai dengan jenjangnya (Alamsyah, 2017).

Setiap tahunnya Universitas Mercu Buana Yogyakarta khususnya di semester akhir mahasiswa melaksanakan bimbingan tugas akhir, namun terdapat beberapa kendala dalam hal pengelolaan data administrasi tugas akhir yang saat ini masih terkomputerisasi secara sederhana, dimana data tugas akhir masih dikelola dalam format excel (\*.xlsx), DVD, dan word (\*.docx), karena pengolahan data yang belum tersistem dan terintegrasi dengan database sehingga membuat dosen sulit menentukan apakah judul tersebut pernah diajukan atau belum, karena masih banyak terjadi mahasiswa yang menggunakan judul skripsi yang sudah pernah digunakan dari angkatan-angkatan yang sebelumnya (Anggadini, 2013).

Dengan adanya permasalahan tersebut, banyak terjadi kasus yang disebut plagiat atau penggunaan judul skripsi yang sudah ada dan digunakan kembali (*plagiarism*), beberapa mahasiswa yang tidak melaporkan progress pengerjaan tugas akhir atau tidak melakukan bimbingan secara rutin, maka diperlukan sebuah sistem informasi yang dapat menyelesaikan masalah tersebut.






Sistem ini diharapkan dapat menyelesaikan permasalahan yang terjadi pada proses pelaksanaan dan manajemen tugas akhir, mengurangi adanya plagiat (*plagiarism*). Untuk itu berdasarkan uraian di atas, maka peneliti dalam hal ini mengambil judul “Deteksi *plagiarism* pada tugas akhir menggunakan metode *rabin-karp* berbasis web” dengan harapan sistem yang dibuat dapat menyelesaikan masalah dengan tepat dan memberikan kemudahan pelaksanaan tugas akhir.

## 1.2 Perumusan Masalah.

Berdasarkan uraian latar belakang, maka perumusan masalah dalam penelitian ini adalah :

1. Bagaimana membuat sebuah sistem yang mampu mendeteksi *plagiarism* terhadap sebuah judul skripsi ?
2. Bagaimana menentukan sebuah skripsi termasuk dalam hasil *plagiarism* atau bukan ?
3. Bagaimana sistem yang diterapkan mengenali kata-kata yang dimiliki kesamaan makna ?
4. Bagaimana menerapkan algoritma *Rabin-karp* yang akan digunakan dalam pengembangan aplikasi pendeteksian plagiat ?

## 1.3 Tujuan Penelitian.

Penelitian ini bertujuan untuk membangun sebuah sistem informasi  yang mampu mendeteksi plagiat (*plagiarism*), pada tugas akhir menggunakan metode *Rabin-karp* berbasis web yang dapat mengurangi terjadinya plagiat atau kesamaan judul skripsi, membantu dosen dalam memberikan keputusan dan mempermudah pelaksanaan tugas akhir, mulai dari proses pengajuan judul dan proposal skripsi hingga pelaksanaan ujian atau sidang skripsi, serta mempermudah proses bimbingan yang bisa dilakukan dimanapun dan kapanpun dengan adanya fasilitas live chat yang ada di dalam sistem

#### **1.4 Manfaat Penelitian.**

Manfaat dari penelitian pembangunan deteksi *plagiarism* pada tugas akhir menggunakan metode *Rabin-Karp* berbasis web ini, yakni memudahkan mahasiswa dalam menyelesaikan tugas akhirnya, memudahkan dosen untuk mengontrol mahasiswa bimbinganya dan memantau progres, serta memudahkan penyimpanan arsip skripsi secara digital dan aman.

Adapun manfaat penelitian ini secara khusus diantaranya adalah sebagai berikut :

1. Memudahkan mahasiswa mengajukan judul dan proposal tugas akhir.
2. Memudahkan dosen pembimbing dan kepala jurusan untuk mengontrol perkembangan progress pengerjaan tugas akhir mahasiswa bimbingannya.
3. Memudahkan Fakultas dalam melakukan arsip atau dokumentasi skripsi.
4. Mengurangi kemungkinan terjadinya pengulangan judul tugas akhir yang sama.
5. Memudahkan mahasiswa dalam mempersiapkan syarat-syarat dan berkas pendukung yang diperlukan dalam administrasi tugas akhir.
6. Memudahkan semua aktor yang berperan dalam sistem untuk melihat informasi jadwal seminar proposal, seminar hasil dan sidang tugas akhir.

## BAB II

### TINJAUAN PUSTAKA DAN LANDASAN TEORI

#### 2.1 Tinjauan Pustaka.

Tinjauan pustaka memperlihatkan perbedaan penelitian yang dilakukan dengan penelitian-penelitian terdahulu, jurnal, artikel atau buku yang saya gunakan sebagai berikut :

Penelitian yang dilakukan oleh (Imam Nawawi, 2019) yang berjudul **“Deteksi plagiarisme pada dokumen skripsi berdasarkan tingkat kesamaan dengan menggunakan metode Longest Common Subsequence”** menyatakan bahwa Plagiarisme adalah tindakan menyalin, mengambil karangan atau pendapat orang lain tanpa adanya izin tertulis dan menjadikannya seolah-olah pendapatnya sendiri. Hal ini masih menjadi fenomena yang sering terjadi pada instansi akademik atau non akademik. Namun pada jurusan Teknologi Informasi Politeknik Negeri Malang belum ada aplikasi yang dapat digunakan untuk mendeteksi plagiarisme. Berdasarkan permasalahan diatas, maka dibuatlah aplikasi deteksi plagiarisme pada dokumen Tugas Akhir / Skripsi yang bernama Document plagiarism Detection (Doristec) dengan menggunakan metode Longest Common Subsequence (LCS) dengan membuat modifikasi untuk mencapai hasil yang sesuai dengan perancangan. Aplikasi ini dibuat dengan tujuan untuk mengetahui seberapa besar tingkat plagiarisme pada dokumen Tugas Akhir / Skripsi yang nantinya akan diketahui oleh mahasiswa dan panitia Laporan Akhir dan Skripsi. Berdasarkan penelitian yang telah dilakukan, Metode Longest Common Subsequence dapat digunakan untuk deteksi plagiarisme dengan perbandingan dua atau lebih dokumen pembanding, dapat menguji lebih dari satu kalimat dan lebih dari satu kandidat pembanding, serta hasil pengujian akurasi yang dilakukan mendapatkan hasil akurasi yang tinggi. Hal ini dapat menjadi alternatif bagi mahasiswa jurusan Teknologi Informasi Politeknik Negeri Malang dalam melakukan

pengujian terhadap penelitiannya dan bagi panitia Laporan Akhir dan Skripsi dapat melakukan monitoring Laporan Akhir dan Skripsi.

Penelitian yang dilakukan oleh (Ibrahim, 2011) yang berjudul **“Pengembangan sistem informasi monitoring tugas akhir berbasis sms”** menjelaskan berdasarkan pengamatan dan survei di Fasilkom Unsri dalam hal pengajuan proposal tugas akhir mahasiswa, mahasiswa harus melalui beberapa prosedur. Proposal yang telah disetujui oleh dosen pembimbing akan diserahkan ke bagian administrasi jurusan dan selanjutnya diserahkan ke Ketua Jurusan untuk diperiksa. Setelah proposal selesai diperiksa dikembalikan ke administrasi beserta status proposal apakah disetujui/direvisi/ditolak. Jika mahasiswa telah selesai menyelesaikan laporan tugas akhirnya, mahasiswa mendaftar ke administrasi jurusan untuk mengikuti ujian tugas akhir. Kegiatan administrasi jurusan tersebut memiliki beberapa kendala dan terlihat kurang efisien, dan tidak efektif. Hal ini terlihat seperti sistem informasi masih belum dapat melacak posisi proposal saat dalam proses pengajuan hingga selesai, sehingga menyebabkan mahasiswa harus sering ke kampus untuk menanyakan status proposalnya, dan untuk melihat jadwal ujian akhir. Selain itu administrasi harus berkoordinasi terlebih dahulu dengan Ketua Jurusan jika ada mahasiswa yang menanyakan status proposalnya. Akibatnya kinerja administrasi terlihat kurang efektif.

Penelitian yang dilakukan oleh (Satia Suhada, 2017) yang berjudul **“Implementasi algoritma Rabin-Karp dan stemming najief andriani untuk deteksi plagiarisme”** Plagiarisme merupakan pengambilan karya orang lain kemudian menjadikannya nampak seperti karya sendiri. Praktek plagiat sering terjadi di dunia akademis seperti pada dokumen proyek akhir. Untuk meminimalkan tindakan plagiarisme, diperlukan suatu sistem untuk menilai atau mengukur seberapa banyak kemiripan dalam sebuah document seperti tugas akhir dan makalah ilmiah beberapa algoritma telah digunakan dalam proses deteksi plagiarisme namun pada penelitian ini akan

menerapkan . Algoritma Rabin Karp dalam mendeteksi plagiarisme karena algoritma ini terbukti efektif untuk membandingkan patern-patern yang ada pada sebuah essai dengan menggunakan fungsi hashing yang dapat menemukan bentuk-bentuk / pola dalam teks, untuk lebih meningkatkan keakuratan dalam proses penemuan pola pada sebuah teks kemudian digunakan algoritma Steeming Najief Andriani yang dapat menemukan kata-kata yang setara yang memiliki persamaan kata dasar yang sama, selanjutnya kedua algoritma yang digunakan akan diimplemntasikan kepada sebuah aplikasi web untuk pengujian document tugas akhir.

Menurut penelitian (Sulhan, 2015) yang berjudul **“Sistem Monitoring Tugas Akhir Berbasis User Generated Content Pada Program Studi Sistem Informasi Universitas Kanjuruhan Malang”**

Pengerjaan Tugas Akhir diperlukan bimbingan dosen atau konseling agar tugas akhir yang dikerjakan sesuai dengan ketentuan yang ada dan sesuai dengan kaidah-kaidahnya, Permasalahannya adalah tidak adanya waktu ataupun kesulitan mahasiswa dalam menemui dosen pembimbing untuk konsultasi tentang tugas akhir yang sedang dikerjakannya, yang dikarenakan kesibukan dari dosen itu sendiri ataupun ketidak sesuaian jadwal antara dosen dan mahasiswa. Konsep *User Generated Content* (UGC) dikenal dengan sifat interaksi dua arah yang lebih menghadirkan kolaborasi, partisipasi, hal ini yang dimungkinkan untuk dikembangkan suatu sistem yang dapat memberikan solusi yakni perlunya dikembangkan sistem manajemen monitoring Tugas Akhir berbasis website. Dalam penelitian ini telah berhasil dibuat sebuah Perangkat lunak Sistem manajemen monitoring tugas akhir online dengan menerapkan konsep UGC, dan Setelah dilakukan Uji Coba pada perangkat lunak tersebut bahwa semua fasilitas dapat berjalan dengan baik sesuai dengan yang diharapkan dan khususnya penerapan konsep UGC, sehingga pengguna dapat saling komunikasi dengan memberikan banyak masukan atau berbagi pengetahuan untuk kemudahan dan kualitas hasil akhir dari tugas akhir, dan diharapkan penyelesaian tugas akhir selesai tepat pada

waktunya.

Penelitian yang telah dilakukan oleh (Alamsyah, 2017) yang berjudul **“Perbandingan algoritma winnowing dengan algoritma rabin karp untuk mendeteksi plagiarisme pada kemiripan teks judul skripsi”** mengatakan plagiarisme biasanya sering terjadi di dunia Pendidikan, baik itu dalam pengerjaan skripsi atau karya tulis ilmiah, dimana mahasiswa dituntut untuk menyelesaikan skripsinya sebagai syarat kelulusan dan mendapatkan gelar sarjana. Deadline untuk segera menyelesaikan karya tulis ilmiah inilah memicu mereka untuk mencari jalan pintas untuk melakukan copy-paste karya ilmiah atau skripsi orang lain sehingga terjadinya plagiarisme. Untuk membantu mengurangi tingkat plagiarisme telah banyak dibuat sistem pendeteksi plagiarisme untuk memeriksa tingkat kemiripan teks dokumen. Pada umumnya, sistem pendeteksi plagiarisme dikembangkan untuk data teks seperti essay, artikel, jurnal, penelitian dan dokumen teks yang lebih terstruktur seperti source code bahasa pemrograman. Untuk mendeteksi plagiarisme atau kemiripan teks dokumen dengan metode fingerprinting dapat dilakukan dengan perbandingan algoritma yang berkaitan pada bidang text mining misal perbandingan pendekatan Algoritma Rabin Karp, Algoritma Manber dan Algoritma Winnowing.

## 2.2 Landasan Teori.

### 2.2.1 Plagiarisme

Plagiarisme berasal dari kata Latin, “plagiarius”, yang berarti “pencuri”. Plagiarisme didefinisikan sebagai tindakan atau praktik mengambil dan mengumpulkan atau menyampaikan pemikiran, tulisan atau hasil karya orang lain selayaknya hasil karya diri sendiri tanpa persetujuan dari pemilik hasil karya tersebut. Dengan kata lain, mempraktikkan plagiarisme berarti mencuri hasil karya atau kepemilikan intelektual orang lain. Plagiarisme adalah tindakan penyalahgunaan, pencurian/perampasan, penerbitan, pernyataan, atau menyatakan sebagai milik sendiri sebuah pikiran, ide, tulisan, atau ciptaan yang sebenarnya milik orang lain. Sedangkan menurut Kamus

Besar Bahasa Indonesia (KBBI), Plagiat merupakan pengambilan karangan (pendapat, dan sebagainya) orang lain dan menjadikannya seolah-olah karangan (pendapat dan sebagainya) sendiri, misalnya menerbitkan karya tulis orang lain atas nama dirinya sendiri (Ngroho, 2011).

Plagiat dapat dianggap sebagai tindak pidana karena mencuri hak cipta orang lain. Pelaku plagiat disebut sebagai plagiator. Sistem pendeteksi plagiarisme dapat dikembangkan untuk :

1. Data teks seperti essay, artikel, jurnal, penelitian dan sebagainya.
2. Dokumen teks yang lebih terstruktur seperti bahasa pemrograman. Berikut ini adalah hal-hal yang tergolong kedalam tindakan plagiarisme, antara lain :
3. Mengakui tulisan orang lain sebagai tulisan sendiri.
4. Mengakui gagasan orang lain sebagai pemikiran sendiri.
5. Mengakui temuan orang lain sebagai kepunyaan sendiri.
6. Mengakui karya kelompok sebagai kepunyaan atau hasil sendiri.

### **2.2.2 Algoritma Rabin Karp**

*String matching* atau pencocokan *string* adalah subjek yang penting dalam kaitanya dengan *text-processing*. Penggunaan *string matching* mencakup pencarian pola dalam *DNA sequence*, *search engine* internet, menemukan halaman web yang relevan pada *query*, dapat pula dimanfaatkan untuk mendeteksi adanya *plagiarism* karya tulis. Termasuk dalam algoritma *string matching* diantaranya algoritma Naive, algoritma Rabin Karp, algoritma Finite Automation, dan algoritma Knuth Morris Pratt (T. H. Cormen, 2001).

Algoritma Rabin Karp ditemukan oleh Michael O. Rabin dan Richard M. Karp. Algoritma ini menggunakan metode *hash* dalam mencari suatu kata. Teori ini jarang digunakan untuk mencari kata tunggal, namun cukup penting dan sangat efektif bila digunakan untuk

pencarian jamak (Atmopawiro, 2006).

Prinsip kerja dari *algoritma Rabin Karp* ini memberikan atau mempresentasikan setiap karakter dalam bentuk desimal digit (*digit radix-d*)  $\Sigma = \{0, 1, 2, 3, \dots, d\}$ , dimana  $d = |\Sigma|$ . Sehingga didapat masukan *string k* berturut-turut sebagai perwakilan panjang *k* desimal. Karakter *string* 31415 sesuai dengan jumlah desimal 31,415. Kemudian pola *p* di-hash menjadi nilai desimal dan *string* direpresentasikan dengan penjumlahan *digit-digit* angka menggunakan aturan Horner's, misal (Elchison, 2012) :

$$\{ A, B, C, \dots, Z \} \rightarrow \{ 0, 1, 2, \dots, 26 \}$$

$$\bullet \text{ BAN} \rightarrow 1 + 0 + 13 = 14$$

$$\bullet \text{ CARD} \rightarrow 2 + 0 + 17 + 3 = 22$$

Untuk pola yang panjang dan teks yang besar, algoritma ini menggunakan operasi *mod*, setelah dikenai operasi *mod q*, nilainya akan menjadi lebih kecil dari *q*.

Tetapi tidak semua nilai *hash* yang cocok berarti polanya cocok. Hal ini sering terjadi pada beberapa kasus, ini disebut *spurious hits*. Kemungkinan terjadinya diantaranya karena:

- Operasi *mod* terinterferensi oleh keunikan nilai *hash* (nilai *mod q* biasanya dipilih bilangan prima sehingga  $10q$  hanya cocok dengan 1 kata komputer)

$$14 \bmod 13 = 1$$

$$27 \bmod 13 = 1$$

- Informasi hilang setelah penjumlahan

$$\text{BAN} \rightarrow 1 + 0 + 13 = 14 \quad \text{CAM} \rightarrow 2 + 0 + 12 = 14$$

Rumus matematis:

$$t_{s+1} = (d(t_s - T[s+1]h) + T[s+m+1]) \bmod q$$

dimana :

$t_s$  = nilai desimal dengan panjang *m* dari *substring*  $T[s+1 .. s+m]$ , untuk

$$s = 0, 1, \dots, n - m$$

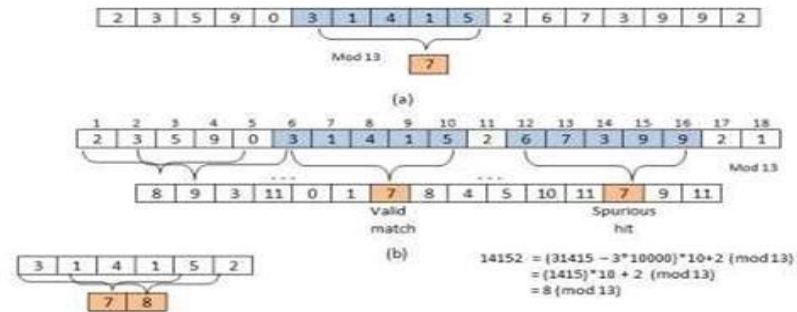
$t_{s+1}$  = nilai desimal selanjutnya yang dihitung dari  $t_s$



$d = \text{radix desimal (bilangan basis 10)}$   $m = \text{panjang pola}$

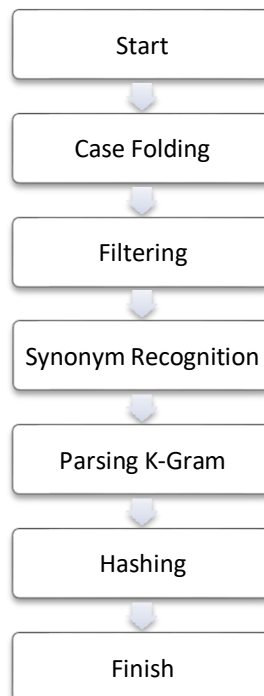
$h = d^{m-1}$   $q = \text{nilai modulo}$

$n = \text{panjang teks}$



### 2.2.3 Preprocessing

Tahap ini melakukan analisis semantik (kebenaran arti) dan sintaktik (kebenaran susunan) teks. Tujuan dari pemrosesan awal adalah untuk mempersiapkan teks menjadi data yang akan mengalami pengolahan lebih lanjut. Operasi yang dapat dilakukan pada tahap ini meliputi part – of – speech (PoS) tagging, menghasilkan parse tree untuk tiap – tiap kalimat, dan pembersihan teks. (Alamsyah, 2017).



#### **2.2.4 Case Folding**

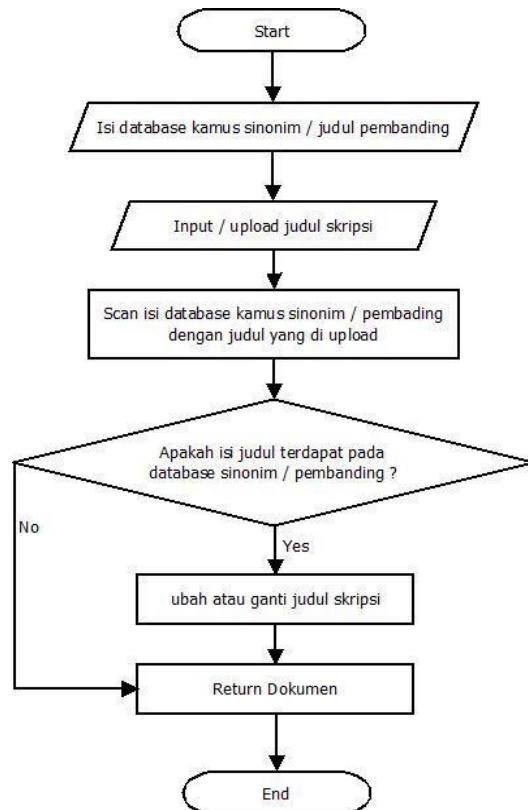
Case folding adalah merupakan tahapan yang mengubah semua huruf dalam dokumen atau isi konten yang terdapat pada suatu web menjadi huruf kecil. Hanya huruf „a“ sampai dengan „z“ yang diterima. Karakter selain huruf dihilangkan dan akan dilakukan proses penghapusan. Pada penelitian ini case folding yang diterima adalah huruf „a“ sampai dengan „z“, selain itu seperti tanda hubung atau tanda baca akan dihapus dan dihilangkan (Putra, 2013).

#### **2.2.5 Filtering**

Filetring adalah tahap mengambil kata – kata penting dari hasil token. Bisa menggunakan algoritma stoplist (membuang kata yang kurang penting) atau wordlist (menyimpan kata penting). Stoplist / stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words. Contoh stopwords adalah “yang”, “dan”, “di”, “dari” dan seterusnya (Langgeni, 2010)

#### **2.2.6 Synonym Recognition**

Synonym Recognition atau pengenalan sinonim adalah pendeteksian plagiarisme melalui pendekatan sinonim. Dalam hal ini dokumen satu dibandingkan dengan dokumen lain dengan mendeteksi kata-kata yang mengandung sinonim sehingga tingkat kesamaan dapat dikatakan lebih akurat.



Pada proses pencocokan sinonim, kaitannya antara isi dokumen dengan kamus sinonim yang terdapat pada database. Jika kata tidak tersedia pada kamus sinonim, maka tidak akan melakukan proses synonym recognition. Seluruh isi dokumen akan discan dan dicocokkan dengan kata yang ada dalam kamus, kemudian akan diubah berdasarkan isi kamus sinonim tersebut (Pratama, 2013).

### 2.2.7 Hashing

Hashing adalah suatu cara untuk mentransformasikan sebuah string menjadi suatu nilai yang unik dengan panjang tertentu (fixed-length) yang berfungsi sebagai penanda string tersebut. Hash function atau fungsi hash adalah suatu cara menciptakan fingerprint dari berbagai data masukan. Hash function akan mengganti atau akan mentransposekan data tersebut untuk menciptakan fingerprint yang biasa disebut hash value (Firdaus, 2008).

### 2.2.8 Parsing K-grams

K-grams adalah rangkaian terms dengan panjang K. Kebanyakan yang digunakan sebagai terms adalah kata. Kgram merupakan sebuah metode yang diaplikasikan untuk pembangkitan kata atau karakter. Metode K-grams ini digunakan untuk mengambil potonganpotongan karakter huruf sejumlah k dari sebuah kata yang secara kontinuitas dibaca dari teks sumber hingga akhir dari dokumen.

Berikut ini adalah contoh K-grams dengan  $k=5$  : Teks : “A do run run run ,a do run run” kemudian dilakukan penghilangan spasi “adorunrunrunadorunrun” sehingga dihasilkan rangkaian 5-grams yang diturunkan dari teks : “adoru dorun orunr runru unrun nrunr runru unrun nruna runad unado nador adoru dorun orunr runru unrun” (Schleimer, 2003).

## **BAB III METODOLOGI**

Bab ini memaparkan tentang metodologi penelitian yang terdiri dari desain penelitian, alat dan bahan penelitian, serta metode penelitian yang digunakan.

### **3.1 Bahan Penelitian**

Objek penelitian ini adalah Universitas Mercu Buana Yogyakarta. Data kriteria yang akan digunakan sebagai parameter penting untuk perhitungan dalam deteksi plagiarisme adalah judul skripsi dari beberapa tahun sebelumnya.

### **3.2 Alat Penelitian**

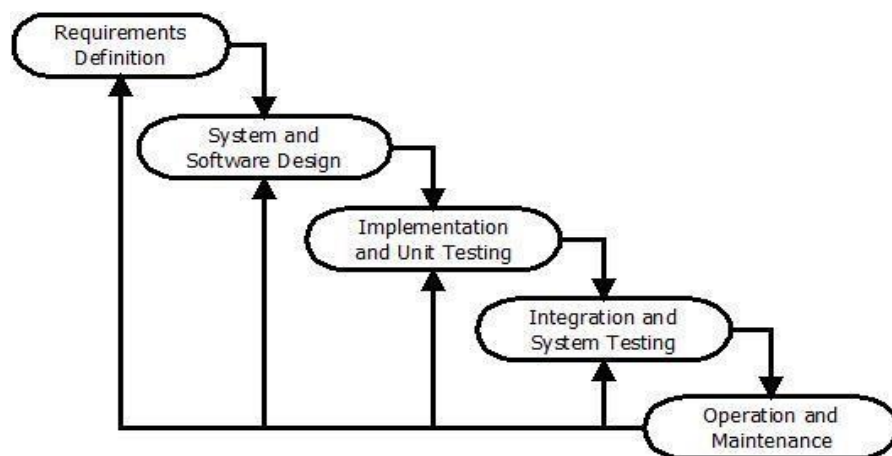
Berikut ini adalah pembahasan lebih detail mengenai alat dan bahas penelitian yang digunakan pada penelitian ini.

- a. Perangkat keras yang digunakan berupa laptop dengan spesifikasi
  - *Processor* Intel(R) Celeron(R) CPU N3350 @ 1.10GHz  
1.10 GHz
  - Harddisk Drive 500 GB
  - SSD Drive 120 GB
- b. Perangkat Lunak yang digunakan untuk perancangan aplikasi berupa
  - Visual studio code 64-1.46.0
  - Sistem Operasi Windows 10 Pro 64 bit
  - PHP 7
  - XAMPP v3.2.4

### 3.3 Jalan Penelitian

#### 3.3.1 Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak ini dilakukan dengan metode waterfall. Model SDLC air terjun (waterfall) sering juga disebut model sekuensial linier (sequential linier). Peneliti menggunakan metode modern waterfall agar jika suatu saat ada kesalahan pada salah satu tahap, bisa dikembalikan ke tahap sebelumnya. Berikut pengertian dari tahap-tahap pada model waterfall.



##### 1. *Requirements Analysis and Definition* (Analisis)

Analisis adalah tahap menentukan aplikasi atau software seperti apakah yang akan dibuat. Analisis merupakan tahapan penetapan fitur, ruang lingkup, tujuan dan batasan yang dimiliki oleh perangkat lunak. Semua hal tersebut akan ditetapkan secara rinci dan berfungsi sebagai spesifikasi perangkat lunak. Dalam penelitian ini proses analisis yang dilakukan mencakup analisis terhadap data yang digunakan dan batasannya yaitu perangkat lunak hanya dapat memproses data berbentuk pdf dan keluaran perangkat lunak yaitu persentase kesamaan dokumen input dengan dokumen pada database.

##### 2. *System and Software Design* (Perancangan)

Dalam tahapan ini dilakukan perancangan lebih detail yang dikembangkan dari hasil analisis. Pada penelitian ini, dilakukan

perancangan antarmuka atau mock-up dari perangkat lunak. Pada tahapan ini juga dibahas gambaran umum mengenai alur dan proses perangkat lunak yang dibuat.

### 3. *Implementation and Unit Testing (Coding)*

*Coding* adalah tahap proses implementasi dari desain, dalam tahapan ini, hasil dari desain perangkat lunak akan direalisasikan sebagai satu set program atau unit program. Setiap unit akan diuji apakah sudah memenuhi spesifikasinya.

### 4. *Integration and System Testing (Testing)*

Perangkat lunak yang dibuat akan memasuki tahap pengujian atau testing. Pengujian yang dilakukan meliputi pengujian fungsi-fungsi yang terdapat pada perangkat lunak dan juga pengujian terhadap data yang digunakan, yaitu data dokumen yang telah di plagiat.

### 5. *Operation and Maintenance (Pemeliharaan)*

Dalam tahapan ini, perangkat lunak mulai digunakan. Selain itu juga memperbaiki error yang tidak ditemukan pada tahap pembuatan. Dalam tahap ini juga dilakukan pengembangan perangkat lunak seperti penambahan fitur dan fungsi baru.

## **3.3.2 Perancangan Proses**

Sistem deteksi plagiat secara umum dirancang untuk dapat mendeteksi kemiripan isi Judul skripsi, yang dimungkinkan kemiripan ini adalah hasil plagiat. Input sistem diperoleh dari judul yang di input oleh mahasiswa atau user. Judul yang di input otomatis akan tersimpan dalam database sistem.

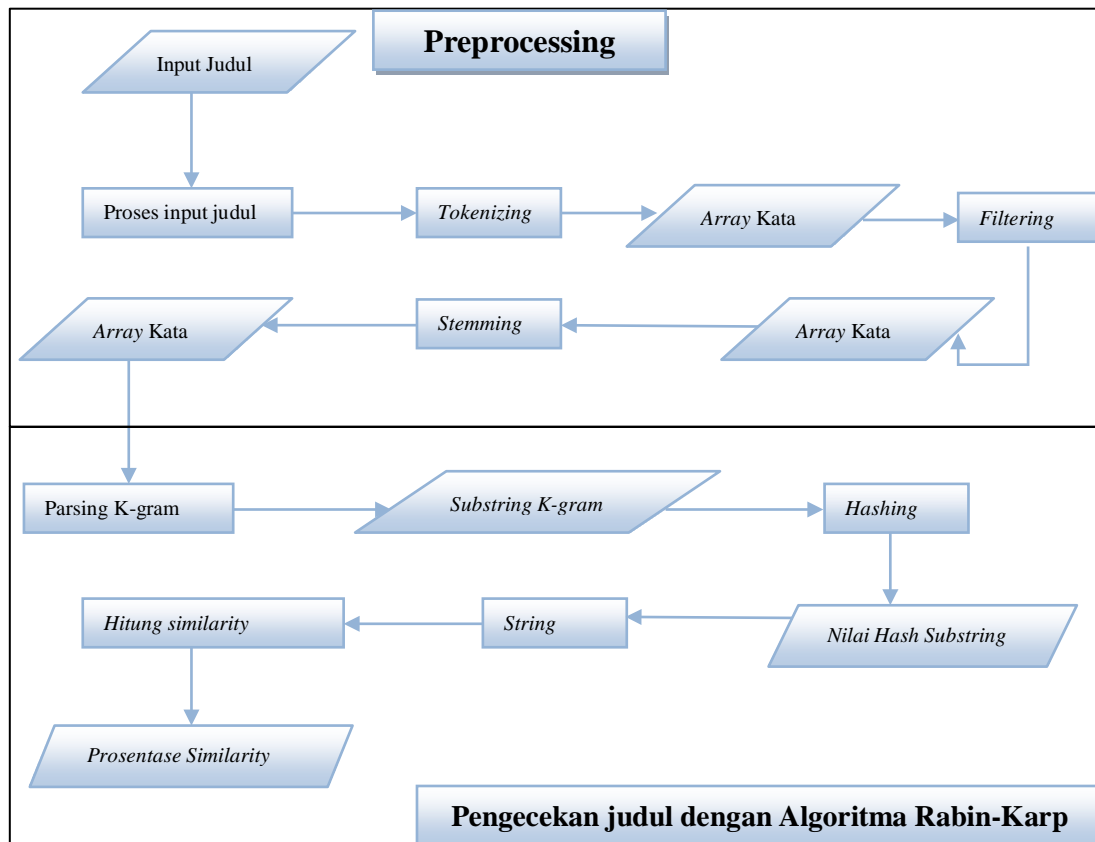


Selanjutnya judul akan melewati tahap *preprocessing* menggunakan *text mining*. Tahap ini terdiri dari *tokenizing* (memecah kalimat menjadi potongan kata sekaligus mengubah isi judul menjadi huruf kecil/*lowercase*), *filtering* (membuang *stopword*/kata yang tidak deskriptif), dan *stemming* (mengembalikan setiap kata ke bentuk dasarnya).

Setelah itu, dokumen hasil *preprocess* akan melewati tahap *processing* menggunakan algoritma Rabin Karp. Tahap ini adalah tahap pencocokan dokumen/judul. Tahap ini terdiri dari *parshing K-gram* (memecah *string* ke dalam potongan *substring* sebanyak *k*), *hashing* (mengubah *substring k* ke dalam nilai *hash*), dan *string matching* (pencocokan *string* hasil *hashing*).

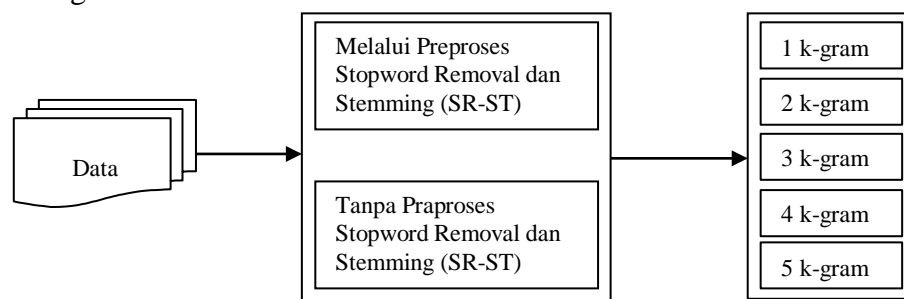
Hasil pencocokan *string* kemudian dihitung nilai *similarity* nya menggunakan *Dice's Similarity coefficient*, yang kemudian nilai tersebut diubah ke dalam bentuk prosentase *similarity*.





### 3.3.3 Skenario Eksperimen

Eksperimen dilakukan dengan melakukan pendeteksian plagiarisme dalam beberapa skenario. Skenario eksperimen ditentukan dengan mengatur praproses yang digunakan dan besar K-Gram yang digunakan.



#### 1. Praproses

Sebelum memasuki proses deteksi plagiarisme dilakukan empat tahap praproses, yaitu case folding,

tokenize, stopword removal dan stemming. Untuk proses deteksi plagiarisme sebenarnya cukup dengan praproses case folding dan tokenize, sehingga praproses stopword removal dan stemming hanya menjadi praproses tambahan agar mencapai hasil yang optimal. Pada eksperimen ini akan dibandingkan skenario eksperimen tanpa menggunakan praproses stopword removal dan stemming (SR-ST) dan skenario eksperimen dengan menggunakan praproses stopword removal dan stemming (SR-ST) untuk mengetahui pengaruh penambahan praproses stopword removal dan stemming pada hasil eksperimen.

## 2. K-Gram

K-gram adalah rangkaian terms dengan panjang K. Kebanyakan yang digunakan sebagai terms adalah kata. K-gram merupakan sebuah metode yang diaplikasikan untuk pembangkitan kata atau karakter. Metode K-gram digunakan untuk mengambil potongan-potongan karakter huruf sejumlah k dari sebuah kata yang secara kontinuitas dibaca dari teks sumber hingga akhir dari dokumen. Contoh penggunaan dari K-grams, trigram dari Menganalisis dan Menganalisa, yaitu “Men, eng, nga, gan, ana, nal, ali, lis, isi, sis”, dan “Men, eng, nga, gan, ana, nal, ali, lis, isa”. Sehingga dapat disimpulkan bahwa tigram dari kedua kata tersebut yang memiliki bentuk yang sama yaitu “Men, eng, nga, gan, ana, nal, ali, lis”.

## **BAB IV**

### **ANALISIS DAN PEMBAHASAN**

#### **4.1 Hasil Penelitian.**

#### **4.2 Analisis dan Pembahasan.**

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan.**

#### **5.2 Saran.**

## DAFTAR PUSTAKA

- Ambler, Scott. 2002. *Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process*. New York: John Wiley & Sons, Inc.
- Gutkowski, Michal; Wojciechowski, Jaroslaw; Sakowicz, Bartosz; Napieralski, Andrzej. 2007. *Thesis Management Supporting System based on J2EE Platform*. IEEE Symposium (Editors). Proceedings of International Conference on The Experience of Designing and Applications of CAD Systems in Microelectronics. Ukraine:Polyana.
- Li Li; Ping Li; Qing Liu; Jian Zhang; Zhongmin Wang; Jungang Han. 2007. *WebUPMS: A Web-based Undergraduate Project Management System*. IEEE Symposium (Editors).Proceeding of Information Technologies and Applications in Education. China: Kunming University.
- Marthasari, Gita Indah. 2009. *Prototipe Sistem Manajemen Tugas Akhir Berbasis Web Menggunakan Teknologi Java Server Faces (JSF)*. Lembaga Penelitian Univesitas Muhamadiyah Malang.
- Pengarang pertama. Tahun. *Teori Pertama*. Kota Penerbit.
- Penerbit. Pengarang kedua. Tahun. *Teori Kedua*. Kota
- Penerbit. Penerbit.
- Pengarang ketiga. Tahun. *Teori Ketiga*. Kota Penerbit.
- Penerbit. Pengarang ke-n. Tahun. *Teori Ke-n*. Kota
- Penerbit. Penerbit.

## LAMPIRAN

### Lampiran A Biodata Peneliti

Nama Lengkap : \_\_\_\_\_  
NIM : \_\_\_\_\_  
Alamat Asal : \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Alamat di Yogyakarta : \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

No. HP. : \_\_\_\_\_  
e-Mail : \_\_\_\_\_  
Website : \_\_\_\_\_

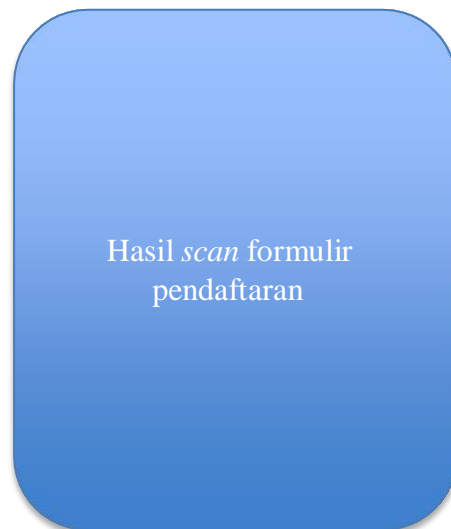
### **Lampiran B *Scan* Dokumen Bimbingan Skripsi**

Dokumen bimbingan skripsi yang dilampirkan berbentuk hasil scan yang telah terisi lengkap berikut tanda tangan dosen pembimbing.

### Lampiran C Artifak/Data Penelitian

Isi yang dilampirkan sebaiknya hasil *scan* dari bentuk fisik yang nyata. Misalnya pada instrumen pendaftaran terdapat 2 formulir yaitu formulir pendaftaran dan surat pernyataan kesanggupan pembayaran administrasi.

Format penamaan lampiran hasil scan seperti pada penamaan gambar, yaitu terletak di bawah, spasi 1, ukuran tulisan 9pt, Times New Roman. Contoh:



**Gambar C. 1 Scan Formulir Pendaftaran**



## Lampiran D Kode Program

Berikut adalah kode program untuk fitur.....

```
1 <?php
2 class proposal extends controller
3 {
4     function proposal()
5     {
6         parent::Controller();
7         $this->security_model->login_mhs_check();
8     }
9
10    function create()
11    {
12        if($_SERVER['REQUEST_METHOD'] == "POST")
13        {
14            $this->form_validation->set_rules('proposal_tema', 'Tema / Topik Tesis', 'trim|xss_clean');
15
16            if ($this->form_validation->run() == FALSE)
17            {
18                $data['err'] = error(validation_errors());
19                $tmp['content'] = $this->load->view("mahasiswa/proposal/create", $data, TRUE);
20            }
21            else
22            {
23                $file_size = $_FILES["image"]["size"];
24                $new_image = $_FILES["image"]["name"];
25
26                $this->proposal_lib->add($mhs_id, $new_image, $file_size, $this->session->userdata('username_mahasiswa'), $ambil_tema_dari);
27            }
28
29            $data['judul'] = "Form Pengajuan Proposal Tesis";
30            // load template
31            $tmp['content'] = $this->load->view("mahasiswa/proposal/create", $data, TRUE);
32            $this->load->view("template", $tmp);
33        }
34    }
35 }
36
37 ?>
```

Gambar D. 1 Kode Program Form Pendaftaran

```
1 <?php
2 class proposal extends controller
3 {
4     function proposal()
5     {
6         parent::Controller();
7         $this->security_model->login_mhs_check();
8     }
9
10    function create()
11    {
12        if($_SERVER['REQUEST_METHOD'] == "POST")
13        {
14            $this->form_validation->set_rules('proposal_tema', 'Tema / Topik Tesis', 'trim|xss_clean');
15
16            if ($this->form_validation->run() == FALSE)
17            {
18                $data['err'] = error(validation_errors());
19                $tmp['content'] = $this->load->view("mahasiswa/proposal/create", $data, TRUE);
20            }
21            else
22            {
23                $file_size = $_FILES["image"]["size"];
24                $new_image = $_FILES["image"]["name"];
25
26                $this->proposal_lib->add($mhs_id, $new_image, $file_size, $this->session->userdata('username_mahasiswa'), $ambil_tema_dari);
27            }
28
29            $data['judul'] = "Form Pengajuan Proposal Tesis";
30            // load template
31            $tmp['content'] = $this->load->view("mahasiswa/proposal/create", $data, TRUE);
32            $this->load->view("template", $tmp);
33        }
34    }
35 }
36
37 ?>
```

Gambar D. 2 Kode Program Proses Penyimpanan Data ke Database.