



**PENERAPAN METODE AGGLOMERATIVE HIERARCHICAL  
CLUSTERING UNTUK KLASIFIKASIH DOKUMEN SKRIPSI**

**ELAN DINATA**

**141420298**

**Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana  
Komputer di Universitas Bina Darma**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BINADARMA  
2020**

**HALAMAN PENGESAHAN**

**PENERAPAN METODE *AGGLOMERATIVE HIERARCHICAL CLUSTERING* UNTUK KLASIFIKASIH DOKUMEN SKRIPSI**

**ELAN DINATA  
141420298**

**Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana  
Komputer di Universitas Bina Darma**

**Pembimbing**



**Hadi Syaputra, M.Kom.**

**Palembang, Maret 2020  
Fakultas Ilmu Komputer  
Universitas Bina Darma  
Dekan**


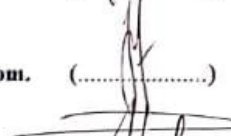
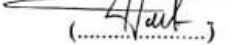


**Dedy Syamsuar, S.Kom., M.I.T., Ph.D.**

## HALAMAN PERSETUJUAN

Skripsi Berjudul "PENERAPAN METEDO *AGGLOMERATIVE HIERARCHICAL CLUSTERING* UNTUK KLASIFIKASIH DOKUMEN SKRIPSI" Oleh "ELAN DINATA" telah dipertahankan didepan komisi penguji pada hari jumat tanggal 24 Februari 2020.

### Komisi Penguji

- |                      |                               |  |
|----------------------|-------------------------------|--|
| 1. Ketua Tim Penguji | Hadi Syaputra, M.Kom.         | <br>(.....)   |
| 2. Anggota Penguji   | A. Haidar Mirza, S.T., M.Kom. | <br>(.....)  |
| 3. Anggota Penguji   | Hutrianto, M.M., M.Kom.       | <br>(.....) |

Mengetahui  
Program Studi Informatika  
Fakultas Ilmu Komputer  
Universitas Bina Darma

Universitas **Bina Darma**  
Fakultas Ilmu Komputer  
Ketua Program Studi  
  
Dr. Widya Cholil, S.KOM, M.I.T.

## PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Elan Dinata

Nim : 141420298

Dengan ini menyatakan bahwa :

1. Skripsi ini adalah asli dan belum pernah di ajukan untuk mendapatkan gelar sarjana di Universitas Bina Darma atau perguruan tinggi lain;
2. Skripsi ini murni gagasan, rumusan dan penelitian saya sendiri dengan arahan tim pembimbing;
3. Di dalam Skripsi ini tidak terdapat karya atau pendapat yang telah di tulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas di kutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan;
4. Saya bersedia Skripsi yang saya hasilkan di cek keasliannya menggunakan plagiarism checker serta di unggah ke internet, sehingga dapat di akses public secara daring;
5. Surat Pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam Pernyataan ini, saya bersedia menerima sanksi sesuai dengan peraturan dan perundang-undangan yang berlaku.

Demikianlah Surat Pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 10 Maret 2020

Yang membuat pernyataan,



Elan Dinata

141420298

## MOTTO DAN PERSEMBAHAN

*Motto :*

- *“Jangan lupa Sholat”*
- *Jadikanla Sholat dan Sabar sebagai penolongmu”*
- *“Berlari dari masalah hanya akan memperpanjang masalah yang dihadapi”*
- *“Selalu libatkan ALLAH SWT dalam segala urusanmu”*
- *“Dalam segala urusanmu selalu minta doa dan ridho kepada kedua orang tua ”*
- *“Jika tidak tahu cara melakukan sesuatu, maka belajar lah agar bisa”*
- *“Jangan pernah putus asah, berdoa dan minta tolong lah kepada ALLAH SWT dan tetap yakin”*

*Persembahan :*

- *Allah SWT yang telah memberi jalan dan kelancaran serta sehat jasmani dan rohani dalam penulisan skripsi ini.*
- *Ayah dan Ibu yang telah mendukung baik secara moral dan material selama penulisan skripsi ini.*
- *Saudara – saudara saya.*
- *Teman – teman yang telah membantu dalam kesusahan.*
- *Ibu Nyimas Sopiah, M.M, M.Kom yang telah memberikan bimbingan selama proses penulisan skripsi*
- *Ibu Dr.Widya Cholil, S.Kom, M.I.T, Ketua Program Studi Informatika.*

## ABSTRAK

Setiap tahunnya Fakultas Teknologi Informatika Bina Darma meluluskan mahasiswa dengan penelitian skripsi yang beragam. Setiap tahunnya jumlah data skripsi selalu bertambah. Semakin bertambahnya penelitian skripsi dengan mata kuliah terbatas menyebabkan semakin banyak pula mahasiswa yang mengambil penelitian yang mirip antara judul dan tema dengan penelitian sebelumnya. Pengklasifikasi data penelitian yang umumnya berbentuk teks dapat dilakukan dengan *text mining*. Tujuan dari *text mining* adalah untuk mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. Terdapat beberapa metode *text mining* salah satunya adalah *clustering*. *Clustering* adalah suatu metode analisa data untuk memecahkan masalah pengelompokan data. Kategori algoritma *clustering* yang banyak dikenal adalah *Hierarchical Agglomerative*. Pada algoritma *clustering*, data akan dikelompokkan menjadi cluster-cluster berdasarkan kemiripan satu data dengan yang lain. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota satu cluster dan meminimumkan kesamaan antar anggota cluster yang berbeda. *Hierarchical Clustering* adalah salah satu algoritma *clustering* yang dapat digunakan untuk meng-cluster dokumen (*document clustering*). Hasil keseluruhan dari algoritma *hierarchical clustering* secara grafik dapat digambarkan sebagai tree, yang disebut dengan dendogram. Tree ini secara grafik menggambarkan proses penggabungan dari cluster cluster yang ada, sehingga menghasilkan cluster dengan level yang lebih tinggi.

**Keyword** : Klasifikasi, , *Text Mining*, *Hierarchical Agglomerative*, *Clustering*

## KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Puji dan syukur kehadiran Allah SWT atas berkat, rahmat dan Karunia-Nya lah, Skripsi ini dapat diselesaikan guna untuk memenuhi syarat dalam menyelesaikan tugas akhir dan bisa lulus dari perkuliaan agar bisa memperoleh gelar sarjana. Disadari penulisan skripsi ini masih jauh untuk bisa dikatakan sempurna, telah dilakukan revisi dan perbaikan lebih dari satu kali agar akhirnya bisa diterima benar, hal tersebut dikarenakan penulis memiliki kemampuan yang terbatas, oleh karena itulah dibutuhkan bimbingan agar bisa menyelesaikan penulisan skripsi ini.

Pada kesempatan yang baik ini, penulis ingin menyampaikan rasa syukur dan ucapan terima kasih kepada pihak – pihak yang telah memberikan dukungan, bimbingan dan nasihat – nasihat selama penulisan skripsi berlangsung. Penulis mengucapkan terima kasih kepada:

1. Dr. Sunda Ariana, M.Pd., M.M. Selaku Rektor Universtas Bina Darma Palembang.
2. Dedy Syamsuar, S.Kom., M.I.T. Selaku Dekan Fakultas Ilmu Komputer.
3. Dr. Widya Cholil, S.Kom.,M.I.T. Selaku Ketua Program Studi Informatika.
4. Hadi Syaputra, M.Kom. Selaku Pembimbing Utama yang telah memberikan arahan dan dukungan dalam penulisan skripsi.
5. Bapak dan Ibu saya yang telah memberikan dukungan moril dan materil kepada saya selama proses penulisan skripsi. Dan juga teman – teman dan saudara saya.

Palembang, Februari 2020  
Penulis

Elan Dinata

## DAFTAR ISI

	<b>Halaman</b>
Cover Dalam.....	i
Halaman Pengesahan.....	ii
Halaman Persetujuan.....	iii
Pernyataan.....	iv
Motto Dan Persembahan.....	v
Abstrak.....	vi
Kata Pengantar.....	vii
Daftar Isi.....	viii
Daftar Gambar.....	xii
Daftar Table.....	xiv
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan dan Manfaat Penelitian.....	5
1.4.1 Tujuan Penelitian .....	5
1.4.2 Manfaat Penelitian .....	5
1.5 Waktu dan Tempat Penelitian .....	5
1.6 Metodologi Penelitian .....	6
1.6.1 Metode Penelitian .....	6



1.6.2	Metode Pengumpulan Data .....	7
1.7	Sistematika Penulisan .....	8
<b>BAB II TINJAUAN PUSTAKA.....</b>		<b>10</b>
2.1	Tinjauan Umum .....	10
2.2	Landasan Teori .....	10
2.2.1	Klasifikasi .....	10
2.2.2	Dokumen Skripsi .....	10
2.2.3	Mahasiswa .....	11
2.2.4	<i>Data Mining</i> .....	12
2.2.5	<i>Text Mining</i> .....	13
2.2.6	<i>Clustering</i> .....	14
2.2.7	<i>Agglomerative Hierarchical Clustering</i> .....	16
2.2.8	RapidMiner .....	17
2.3	Penelitian Terdahulu .....	20
2.3.1	Implementasi Algoritma <i>Hierarchical Clustering</i> Untuk Klasterisasi Dokumen Bahasa Indonesia .....	20
2.3.2	Penerapan <i>Hierarchical Clustering</i> Pada Pengukuran Kohesi Kelas .....	20
<b>BAB III ANALISA DAN PERANCANGAN.....</b>		<b>22</b>
3.1	Metodologi Penelitian .....	22
3.2	Pengumpulan .....	23
3.2.1	Sampel Data .....	23
3.2.2	Alat dan Bahan .....	27
3.2.2.1	Kebutuhan Perangkat Keras ( <i>Hardware</i> ) .....	27
3.2.2.2	Kebutuhan Perangkat Lunak ( <i>Software</i> ) .....	28
3.2.3	Kerangka Kerja Penelitian .....	28

<b>BAB IV HASIL DAN PEMBAHASAN.....</b>	<b>35</b>
4.1 Pengujian ( <i>Examination</i> ) .....	35
4.1.1 Persiapan Data .....	35
4.2 Proses <i>Clustering</i> Dokumen .....	44
4.3 Clustering .....	49
<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>60</b>
5.1 Kesimpulan .....	60
5.2 Saran .....	61
<b>DAFTAR PUSTAKA.....</b>	<b>62</b>
<b>LAMPIRAN.....</b>	<b>64</b>

## DAFTAR GAMBAR

## HALAMAN

<b>Gambar 1.1</b> Kerangka Metode Penelitian .....	7
<b>Gambar 2.2</b> Hubungan Pengelompokan dengan <i>hierarchical</i> .....	15
<b>Gambar 2.3</b> Tampilan Awal <i>RapidMiner</i> .....	19
<b>Gambar 3.1</b> Kerangka Kerja Penelitian .....	29
<b>Gambar 3.2</b> Proses <i>Tokenizing</i> .....	31
<b>Gambar 3.3</b> Proses <i>Filtering</i> .....	31
<b>Gambar 3.4</b> Proses <i>Stemming</i> .....	32
<b>Gambar 3.5</b> Rumus <i>Euclidean</i> .....	33
<b>Gambar 3.6</b> Rumus <i>Linkage</i> .....	<b>33</b>
<b>Gambar 4.1</b> Sampel Judul dan Abstrak Dokumen Skripsi Teknik Informatika Universitas Bina Darma Palembang .....	36
<b>Gambar 4.2</b> Tampilan Utama Aplikasi <i>RapidMiner</i> .....	42
<b>Gambar 4.3</b> Tampilan <i>Marketplace</i> Aplikasi <i>RapidMiner</i> .....	43
<b>Gambar 4.4</b> Proses <i>Create Document</i> .....	44
<b>Gambar 4.5</b> <i>Preprocessing Data</i> .....	45
<b>Gambar 4.6</b> Tampilan Jendela <i>Process Document</i> .....	45
<b>Gambar 4.7</b> Proses <i>Tokenisasi</i> .....	46
<b>Gambar 4.8</b> Proses <i>Filtering</i> .....	47
<b>Gambar 4.9</b> Proses <i>Stemming</i> .....	48
<b>Gambar 4.10</b> Sampel Hasil TF-IDF dengan 371 Dokumen .....	48
<b>Gambar 4.11</b> Sampel Proses <i>Hierarchical Clustering</i> .....	54
<b>Gambar 4.12</b> Dendogram Hasil <i>Hierarchical Clustering</i> .....	55

## DAFTAR TABEL

### HALAMAN

<b>Tabel 3.1</b> Sampel Data Judul Skripsi Teknik Informatika .....	23
<b>Tabel 3.2</b> Spesifikasi Laptop .....	28
<b>Tabel 4.1</b> Sampel Judul dan Abstrak Setelah dilakukan Penyortiran.....	37
<b>Tabel 4.2</b> Tabel Data Studi Kasus .....	<b>50</b>
<b>Tabel 4.3</b> Matriks Jarak <i>Euclidean</i> .....	51
<b>Tabel 4.4</b> Matriks Pembaruan Tahap 1 .....	52
<b>Tabel 4.5</b> Matriks Pembaruan Tahap 2 .....	52
<b>Tabel 4.6</b> Matriks Pembaruan Tahap 3 .....	53
<b>Tabel 4.7</b> <i>Cluster Model</i> .....	55
<b>Tabel 4.8</b> Analisa Kata Dokumen yang Sering Tampil .....	57
<b>Tabel 4.9</b> Analisa Kata Dokumen yang Sedikit Tampil .....	58