

**IMPLEMENTASI ALGORITMA *AGGLOMERATIVE
HIERARCHICAL CLUSTERING* (AHC) GUNA
PENGELOMPOKAN DATA *MEDIA MONITORING*
GEOPARK CILETUH**

TUGAS AKHIR



MUTIARA JULIA IFRA

1162001034

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2023**

**IMPLEMENTASI ALGORITMA *AGGLOMERATIVE
HIERARCHICAL CLUSTERING* (AHC) GUNA
PENGELOMPOKAN DATA *MEDIA MONITORING*
GEOPARK CILETUH**

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer



MUTIARA JULIA IFRA

1162001034

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2023**

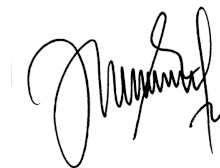
Halaman Pernyataan Orisinalitas

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

Nama : Mutiara Julia Ifra

NIM : 1162001034

Tanda Tangan :



Tanggal : 22 Agustus 2023

Halaman Pengesahan

Tugas akhir ini diajukan oleh :

Nama : Mutiara Julia Ifra

NIM : 1162001034

Program Studi : Teknik Informatika

Fakultas : Teknik dan Ilmu Komputer

Judul Tugas Akhir : Implementasi Algoritma Agglomerative Hierarchical Clustering (AHC) Guna Pengelompokan Data Media Monitoring Geopark Ciletuh

Telah berhasil dipertahankan dihadapan Dewan Penguji sebagai persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie.

DEWAN PENGUJI

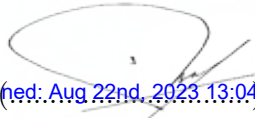
Pembimbing : Guson P. Kuntarto, S.T., M.Sc.


(24/08/2023 15:46 WIB)

Penguji 1 : Irwan Prasetya Gunawan, ST, MEng, PhD.



Penguji 2 : Berkah Iman Santoso, ST, MTI.


e-signed: Aug 22nd, 2023 13:04 WIB

Ditetapkan : Jakarta

Tanggal : 22 Agustus 2023

Ungkapan Terima Kasih

Puji Syukur dipanjatkan kepada Allah SWT atas segala berkah, rahmat dan hidayah-Nya, sehingga terselesaikannya Tugas Akhir ini yang berjudul "Implementasi Algoritma *Agglomerative Hierarchical Clustering* (AHC) Guna Pengelompokan Data *Media Monitoring* Geopark Ciletuh" sebagai salah satu syarat dalam mencapai gelar Sarjana Komputer di Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie.

Penelitian dan penyusunan Tugas Akhir ini tidak akan terwujud tanpa dukungan serta bantuan dari berbagai pihak. Oleh sebab itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada :

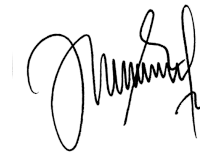
1. Kedua Orang Tua Tercinta, Ibu Hasniarti Pohan dan Bapak Aznir Pasaribu terima kasih untuk doa, kasih sayang, nasihat, pengorbanan dan pengertian yang tidak pernah ada hentinya diberikan untuk penulis sehingga terselesaikannya Tugas Akhir ini.
2. Bapak Guson P. Kuntarto, S.T., M.Sc. selaku dosen Pembimbing Tugas Akhir, yang selalu sedia meluangkan waktu untuk membimbing dengan sabar, memberikan motivasi, nasihat, doa serta dukungan selama proses penelitian dan penyusunan Tugas Akhir ini.
3. Bapak Irwan Prasetya Gunawan, S.T., M.Eng., Ph.D, selaku dosen pembahas seminar proposal dan selaku dosen penguji sidang akhir yang telah memberikan bimbingan, nasihat dan saran yang sangat membantu dalam penyusunan Tugas Akhir ini.
4. Bapak Berkah Iman Santoso, ST, MTI, selaku dosen penguji sidang akhir yang telah memberikan masukan dan perbaikan dalam penyusunan Tugas Akhir ini.
5. Bapak Prof.Dr.Hoga Saragih, ST, MT. selaku dosen Pembimbing Akademik yang telah memberikan semangat di semester awal perkuliahan.

6. Seluruh Bapak dan Ibu dosen pengajar Program Studi Informatika yang telah menginspirasi dan memotivasi penulis dari awal perkuliahan hingga penyusunan Tugas Akhir ini.
7. Teman-teman Informatika 2016, Clara Velita Pranolo, Khairunnisa Savitri, Binya Amary, Refa Nurleana, Elsismone Utari Fitri, Rizchy Hafivah, Rafi Ramadhan Bahari, Fitrah Cahya, Hafiz Kurnia Aji, Pilipus Delevia Vegas, Riski Setiawan dan teman-teman seperjuangan angkatan 2016 lainnya atas kebersamaan, motivasi, semangat, bantuan, dan suka duka selama masa perkuliahan di Universitas Bakrie.
8. Seluruh pihak yang terlibat, baik saudara maupun teman yang telah membantu dan memberikan semangat serta doa dalam penyusunan Tugas Akhir ini.

Semoga Allah SWT membalas segala kebaikan dunia dan akhirat kepada semua pihak yang telah membantu. Peneliti menyadari bahwa penulisan Tugas Akhir ini masih jauh dari kata sempurna, peneliti berharap bahwa penelitian dan Tugas Akhir ini dapat memberikan manfaat dan berguna bagi kalangan pendidikan, khususnya bidang Informatika.

Jakarta, 22 Agustus 2023

Penulis



Mutiara Julia Ifra

Halaman Pernyataan Persetujuan Publikasi

Sebagai civitas akademik Universitas Bakrie, saya yang bertanda tangan dibawah ini :

Nama : Mutiara Julia Ifra
NIM : 1162001034
Program Studi : Teknik Informatika
Fakultas : Teknik dan Ilmu Komputer

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul :

Implementasi Algoritma *Agglomerative Hierarchical Clustering (AHC)* Guna Pengelompokan Data *Media Monitoring* Geopark Ciletuh

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Tanggal : 22 Agustus 2023

Yang Menyatakan,



Mutiara Julia Ifra

IMPLEMENTASI ALGORITMA *AGGLOMERATIVE HIERARCHICAL CLUSTERING* (AHC) GUNA PENGELOMPOKAN DATA *MEDIA MONITORING* GEOPARK CILETUH

Mutiara Julia Ifra

ABSTRAK

Setiap perusahaan atau organisasi membutuhkan kegiatan *media monitoring* dalam hal membangun, mengembangkan dan mempertahankan reputasi citra perusahaan. Dalam melakukan kegiatan *media monitoring*, Wisata Geopark Ciletuh memiliki praktisi *Public Relation* (PR) yang bertugas untuk melaksanakan kegiatan *media monitoring*. Praktisi *Public Relation* (PR) Geopark Ciletuh melakukan identifikasi dan analisis berita secara manual. Penelitian ini mengusulkan dua algoritma *machine learning* untuk melakukan analisis berita dalam *clustering* menjadi lebih efektif dan efisien. Algoritma yang digunakan yaitu *Agglomerative Hierarchical Clustering* (AHC) dan *K-Means* serta *Euclidean Distance* sebagai metode untuk mengukur kemiripan antar teks dalam proses *clustering* terhadap tema berita dan isi berita. AHC diimplementasikan dengan menggunakan tiga parameter yaitu *single linkage*, *average linkage* dan *complete linkage*. Hasil penelitian dengan algoritma AHC dengan parameter jarak *single linkage* menghasilkan *cluster* sebanyak dua *cluster*, *average linkage* menghasilkan *cluster* sebanyak dua *cluster* dan *complete linkage* menghasilkan *cluster* sebanyak dua *cluster*. Sedangkan algoritma *K-means* menghasilkan *cluster* sebanyak tiga *cluster*. Dalam menentukan metode pengelompokan paling baik dari akurasi yang didapatkan. Algoritma AHC pada parameter *average linkage* dengan *euclidean distance* memperoleh rata-rata tertinggi dengan nilai *silhouette* sebesar 0.65. Sedangkan Algoritma *K-means* dengan *euclidean distance* memperoleh rata-rata tertinggi dengan nilai *silhouette* sebesar 0.57.

Kata Kunci : *Agglomerative Hierarchical Clustering, Single Linkage, Average Linkage, Complete Linkage, K-Means, Euclidean Distance, Silhouette Coefficient*

IMPLEMENTATION OF AGGLOMERATIVE HIERARCHICAL CLUSTERING (AHC) FOR MEDIA MONITORING DATA CLUSTERING GEOPARK CILETUH

Mutiara Julia Ifra

ABSTRACT

Every company or organization needs media monitoring activities in terms of building, developing and maintaining a company's image reputation. In carrying out media monitoring activities, Geopark Ciletuh tourism has a Public Relations (PR) practitioner whose job is to carry out media monitoring activities. Geopark Ciletuh Public Relation (PR) practitioners carry out identification and analysis of news manually. This study proposes two machine learning algorithms to make news analysis in clustering more effective and efficient. The algorithms used are Agglomerative Hierarchical Clustering (AHC) and K-Means and Euclidean Distance as a method for measuring similarities between texts in the clustering process for news themes and news content. AHC is implemented using three parameters, namely single linkage, average linkage and complete linkage. The results of the research using the AHC algorithm with the parameter distance of a single linkage produce clusters of two clusters, average linkage produces clusters of two clusters and complete linkage produces clusters of two clusters. While the K-means algorithm produces clusters of three clusters. In determining the best clustering method from the accuracy obtained. The AHC algorithm on the average linkage parameter with the euclidean distance obtains the highest average with a silhouette value of 0.65. While the K-means algorithm with euclidean distance gets the highest average with a silhouette value of 0.57.

Keywords : *Agglomerative Hierarchical Clustering, Single Linkage, Average Linkage, Complete Linkage, K-Means, Euclidean Distance, Silhouette Coefficient*

Daftar Isi

Halaman Pernyataan Orisinalitas	i
Halaman Pengesahan	ii
Ungkapan Terima Kasih	iii
Halaman Pernyataan Persetujuan Publikasi	v
Abstrak	vi
Abstract	vii
Daftar Isi	vii
Daftar Tabel	xi
Daftar Gambar	xiii
Daftar Skrip	xv
Daftar Singkatan	xvii
I Pendahuluan	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	7
1.3 Tujuan Penelitian	7
1.4 Manfaat Penelitian	7
1.5 Ruang Lingkup Penelitian	8

1.6	Kontribusi Penelitian	8
1.7	Sistematika Penulisan	10
II	Tinjauan Pustaka	11
2.1	Penelitian Terkait	11
2.2	<i>Media Monitoring</i>	16
2.3	<i>Machine Learning</i>	17
2.4	<i>Clustering</i>	18
2.5	<i>Distance Measure</i>	19
2.6	Algoritma <i>Agglomerative Hierarchical Clustering</i> (AHC)	22
2.6.1	<i>Single Linkage</i>	24
2.6.2	<i>Average Linkage</i>	25
2.6.3	<i>Complete Linkage</i>	25
2.7	<i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	26
2.8	Pengukuran Performa Algoritma AHC	28
III	Metodologi Penelitian	31
3.1	Tahapan Penelitian	32
3.1.1	Studi Literatur	33
3.1.2	Merumuskan Tujuan Penelitian	34
3.1.3	Pengumpulan Data	34
3.1.4	Pelaksanaan Penelitian	34
3.1.5	Hasil Implementasi dan Pembahasan	35
3.1.6	Penyusunan Laporan Hasil Penelitian	35
3.2	Kerangka Penelitian	36
3.2.1	Kerangka Kerja	36
3.2.2	Implementasi Algoritma <i>Agglomerative Hierarchical Clustering</i> (AHC)	38
3.2.3	<i>Flowchart Single Linkage</i> Pada Algoritma AHC	40
3.2.4	<i>Flowchart Average Linkage</i> Pada Algoritma AHC	41
3.2.5	<i>Flowchart Complete Linkage</i> Pada Algoritma AHC	41

3.3	Instrumen Penelitian	42
IV	Implementasi dan Hasil Penelitian	43
4.1	Pengumpulan Data	43
4.2	Pengolahan <i>Dataset</i>	44
4.2.1	<i>Filtering</i>	47
4.2.2	<i>Tokenizing</i>	48
4.2.3	<i>Case Folding</i>	48
4.2.4	<i>Remove Stopword</i>	48
4.2.5	<i>Stemming</i>	48
4.3	Pembobotan dengan <i>Term Frequency - Inverse Document Frequency</i> (TF-IDF)	51
4.4	Implementasi <i>Agglomerative Hierarchical Clustering</i> AHC	53
4.4.1	Implementasi <i>Single Linkage</i>	54
4.4.2	Implementasi <i>Average Linkage</i>	55
4.4.3	Implementasi <i>Complete Linkage</i>	56
4.5	Implementasi <i>K-Means</i>	58
4.6	Eksperimen	60
4.6.1	Hasil Uji Coba <i>K-Means</i>	60
4.6.2	Hasil Uji Coba <i>Single Linkage</i>	63
4.6.3	Hasil Uji Coba <i>Average Linkage</i>	65
4.6.4	Hasil Uji Coba <i>Complete Linkage</i>	67
4.7	Pembahasan	70
V	Simpulan dan Saran	72
5.1	Simpulan	72
5.2	Saran	73
A	<i>Raw Dataset</i>	79
B	Hasil Pembobotan <i>TF-IDF</i>	86

C Hasil Anggota <i>Clustering</i> AHC - <i>Single Linkage</i> dengan <i>Euclidean Distance</i>	91
D Hasil Anggota <i>Clustering</i> AHC - <i>Average Linkage</i> dengan <i>Euclidean Distance</i>	95
E Hasil Anggota <i>Clustering</i> AHC - <i>Complete Linkage</i> dengan <i>Euclidean Distance</i>	99
F Source Code	102

Daftar Tabel

2.1	Rangkuman Penelitian Terkait dan Kontribusi	15
4.1	Rincian Jumlah Data Berita	44
4.2	Contoh Potongan Data Yang Digunakan	45
4.3	Tahap <i>Text Preprocessing</i> pada potongan Tema Berita	50
4.4	Nilai <i>Silhouette Coefficient K-Means</i>	63
4.5	Nilai <i>Silhouette Coefficient Single Linkage</i>	64
4.6	Nilai <i>Silhouette Coefficient Average Linkage</i>	67
4.7	Nilai <i>Silhouette Coefficient Complete Linkage</i>	69
4.8	Nilai <i>Silhouette Coefficient</i> Tertinggi setiap parameter jarak dengan <i>Euclidean Distance</i>	70
C.1	Hasil Anggota <i>Clustering - Single Linkage AHC</i>	91
C.2	Hasil Anggota <i>Clustering - Single Linkage AHC</i>	93
C.3	Hasil Anggota <i>Clustering - Single Linkage AHC</i>	94
C.4	Hasil Anggota <i>Clustering - Single Linkage AHC</i>	94
D.1	Hasil Anggota <i>Clustering - Average Linkage AHC</i>	95
D.2	Hasil Anggota <i>Clustering - Average Linkage AHC</i>	97
D.3	Hasil Anggota <i>Clustering - Average Linkage AHC</i>	98
D.4	Hasil Anggota <i>Clustering - Average Linkage AHC</i>	98
E.1	Hasil Anggota <i>Clustering - Complete Linkage AHC</i>	99
E.2	Hasil Anggota <i>Clustering - Complete Linkage AHC</i>	99
E.3	Hasil Anggota <i>Clustering - Complete Linkage AHC</i>	100

E.4 Hasil Anggota <i>Clustering - Complete Linkage AHC</i>	101
--	-----

Daftar Gambar

2.1	Arsitektur penelitian <i>Clustering</i> Geopark Ciletuh	16
2.2	Model Pembelajaran <i>Machine Learning</i> [3]	17
2.3	Contoh Visualisasi Metode <i>Clustering</i> [10, 14]	19
2.4	Struktur Dendogram [3, 13, 32]	22
2.5	Visualisasi <i>Single Linkage</i> [22]	24
2.6	Visualisasi <i>Average Linkage</i> [22]	25
2.7	Visualisasi <i>Complete Linkage</i> [22]	25
3.1	Tahap penelitian Implementasi AHC pada data Geopark Ciletuh	33
3.2	Kerangka kerja penelitian AHC pada data destinasi wisata Geopark Ciletuh	36
3.3	Proses Implementasi AHC dan <i>K-Means</i> pada data Geopark Ciletuh	38
3.4	<i>Flowchart Single Linkage</i> pada Algoritma AHC	40
3.5	<i>Flowchart Average Linkage</i> pada Algoritma AHC	41
3.6	<i>Flowchart Complete Linkage</i> pada Algoritma AHC	42
4.1	Grafik Metode <i>Elbow</i>	60
4.2	Visualisasi <i>K-Means</i>	61
4.3	Grafik <i>Silhouette Score</i> K-Means	62
4.4	Grafik <i>Silhouette Score</i> Single Linkage	64
4.5	Visualisasi Dendrogram <i>Single Linkage</i>	65
4.6	Grafik <i>Silhouette Score</i> Average Linkage	66
4.7	Visualisasi Dendrogram <i>Average Linkage</i>	67

4.8	Grafik <i>Silhouette Score</i> Complete Linkage	68
4.9	Visualisasi Dendrogram <i>Complete Linkage</i>	69

Daftar Skrip

4.1	<i>Script Preprocessing Raw Data</i>	46
4.2	<i>Script</i> untuk mendapatkan bobot TF-IDF	51
4.3	<i>Script</i> Euclidean Distance dan Implementasi Single Linkage . .	54
4.4	<i>Script</i> Euclidean Distance dan Implementasi Average Linkage .	55
4.5	<i>Script</i> Euclidean Distance dan Implementasi Complete Linkage	57
4.6	Visualisasi Dendrogram	58
4.7	<i>Script</i> Implementasi K-Means dengan Library Scikit-Learn . . .	59

Daftar Singkatan

Singkatan	Penjelasan
AHC	: <i>Agglomerative Hierarchical Clustering</i>
ED	: <i>Euclidean Distance</i>
TF-IDF	: <i>Term Frequency Inverse Document Frequency</i>
UGG	: <i>UNESCO Global Geopark</i>
GGN	: <i>Global Geopark Network</i>
EGN	: <i>Europe Geopark Network</i>
UNESCO	: <i>United Nations of Educational, Scientific, and Cultural Organization</i>