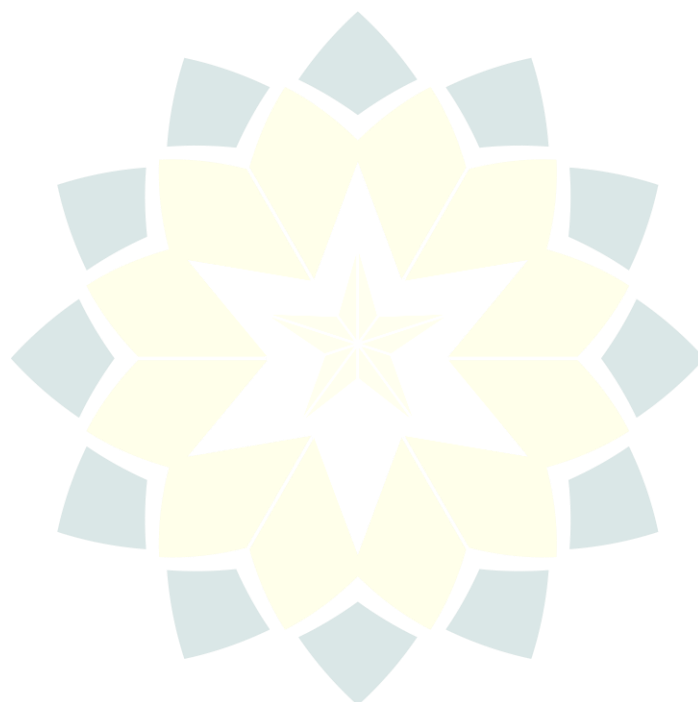


## DAFTAR ISI

<b>LEMBAR PERSETUJUAN</b> .....	i
<b>LEMBAR PENGESAHAN</b> .....	ii
<b>PERNYATAAN KEASLIAN SKRIPSI</b> .....	iii
<b>MOTTO DAN PERSEMBAHAN</b> .....	iv
<b>ABSTRAK</b> .....	v
<b>ABSTRACT</b> .....	vi
<b>KATA PENGANTAR</b> .....	vii
<b>DAFTAR ISI</b> .....	ix
<b>DAFTAR GAMBAR</b> .....	xii
<b>DAFTAR TABEL</b> .....	xiii
<b>DAFTAR SIMBOL</b> .....	xiv
<b>DAFTAR ISTILAH</b> .....	xv
<b>DAFTAR LAMPIRAN</b> .....	xvi
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah .....	4
1.4 Tujuan Penelitian.....	4
1.5 Metode Penelitian .....	5
1.6 Sistematika Penulisan .....	5
<b>BAB II LANDASAN TEORI</b> .....	7
2.1 Hadist .....	7
2.2 <i>Text Mining</i> .....	8
2.3 <i>Clustering</i> .....	9
2.4 <i>Metadata</i> .....	9

2.5	Algoritma <i>COATES</i> .....	11
2.6	Algoritma <i>k-means</i> .....	13
2.7	<i>Text Pre-processing</i> .....	16
2.8	Pembobotan Kata.....	17
2.9	Metode evaluasi <i>Clustering Silhoutte Coefficient</i> .....	18
2.10	Python .....	19
<b>BAB III PROSES <i>CLUSTERING</i> HADIST MENGGUNAKAN</b>		
<b>ALGORITMA <i>COATES</i></b> .....		
3.1	Manajemen Model .....	20
3.2	Tahap Pengambilan Data Hadist .....	22
3.3	Tahap <i>Pre-processing</i> .....	23
3.3.1	<i>Tokenizing</i> .....	23
3.3.2	<i>Case Folding</i> .....	23
3.3.3	<i>Filtering</i> .....	24
3.3.4	<i>Stemming</i> .....	25
3.4	Tahap Pembobotan Kata.....	25
3.5	Tahap <i>Clustering</i> Menggunakan Algoritma <i>COATES</i> .....	27
<b>BAB IV ANALISIS HASIL <i>CLUSTERING</i> HADIST MENGGUNAKAN</b>		
<b>ALGORITMA <i>COATES</i></b> .....		
4.1	<i>Data Set</i> .....	29
4.2	Pengolahan <i>Data Set</i> .....	33
4.3	Hasil Inisialisasi <i>Cluster</i> .....	37
4.4	Hasil <i>Clustering</i> .....	39
<b>BAB V PENUTUP</b> .....		
5.1	Kesimpulan .....	44
5.2	Saran .....	44

<b>DAFTAR PUSTAKA</b> .....	45
<b>LAMPIRAN</b> .....	47



uin

UNIVERSITAS ISLAM NEGERI  
SUNAN GUNUNG DJATI  
BANDUNG

## DAFTAR GAMBAR

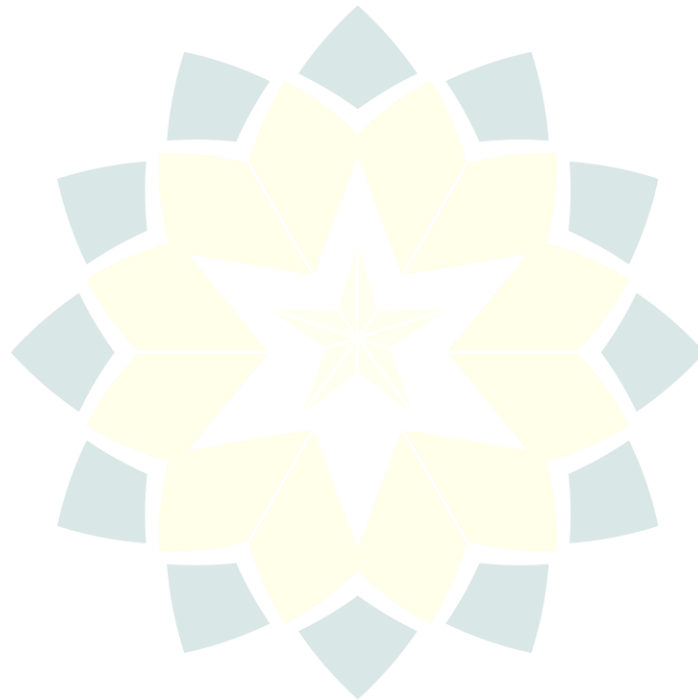
Gambar 2.1. Algoritma <i>k-means</i> .....	15
Gambar 3.1. Alur Penelitian.....	20
Gambar 3.2. Alur Proses <i>Clustering</i> .....	21
Gambar 3.3. Alur <i>Pre-Processing</i> .....	21
Gambar 3.4. Algoritma <i>COATES</i> .....	22
Gambar 3.5. Proses <i>Tokenizing</i> .....	23
Gambar 3.6. Proses <i>Case Folding</i> .....	24
Gambar 3.7. Proses <i>filtering</i> .....	24
Gambar 3.8. Proses <i>Stemming</i> .....	25
Gambar 3.9 Proses Vektor Space Model.....	26
Gambar 3.10 Proses Term Frequency.....	27
Gambar 3.11 Proses Clustering dengan Algoritma <i>COATES</i> .....	28
Gambar 4.1 Hasil evaluasi <i>Clustering</i> dengan <i>Silhouette Coefficient</i> .....	39



UNIVERSITAS ISLAM NEGERI  
SUNAN GUNUNG DJATI  
BANDUNG

## DAFTAR TABEL

Tabel 4.1 Data Hadist dan banyaknya <i>Section</i> pada setiap Bab .....	31
Tabel 4.2 Hasil inialisasi <i>Clustering</i> dengan Algoritma <i>K-means</i> .....	38
Tabel 4.3 Hasil <i>Clustering</i> dengan Algoritma <i>COATES</i> .....	40



uin

UNIVERSITAS ISLAM NEGERI  
SUNAN GUNUNG DJATI  
BANDUNG

## DAFTAR SIMBOL

$k$  : Jumlah *cluster*

$T_i$  : Jumlah dokumen ke- $i$  , dimana  $i = 1, 2, \dots, n$

$X_i$  : Jumlah atribut tambahan ke- $i$  , dimana  $i = 1, 2, \dots, n$

$C_i$  : Inisialisasi *cluster* ke- $i$  , dimana  $i = 1, 2, \dots, k$

$L_i$  : *Centroid* ke- $i$  , dimana  $i = 1, 2, \dots, k$

$\gamma$  : Batas nilai *diskriminatif*

$P_{rj}$  : Peluang keberadaan relatif dokumen

$d$  : dokumen ke- $d$ , dimana  $d = 1, 2, \dots, n$

$i$  : token (kata) ke- $i$ , dimana  $i = 1, 2, \dots, z$

$W$  : Bobot dari setiap *term*.

$Tf$  : Jumlah dari setiap *term*.

$Idf$  : *Inverse Document frequency* dari setiap *term*.

$G_r$  : *Giny indeks* dengan atribut  $r$

$x_j$  : bobot variabel data ke- $j$ , dimana  $j = 1, 2, \dots, z$



uin

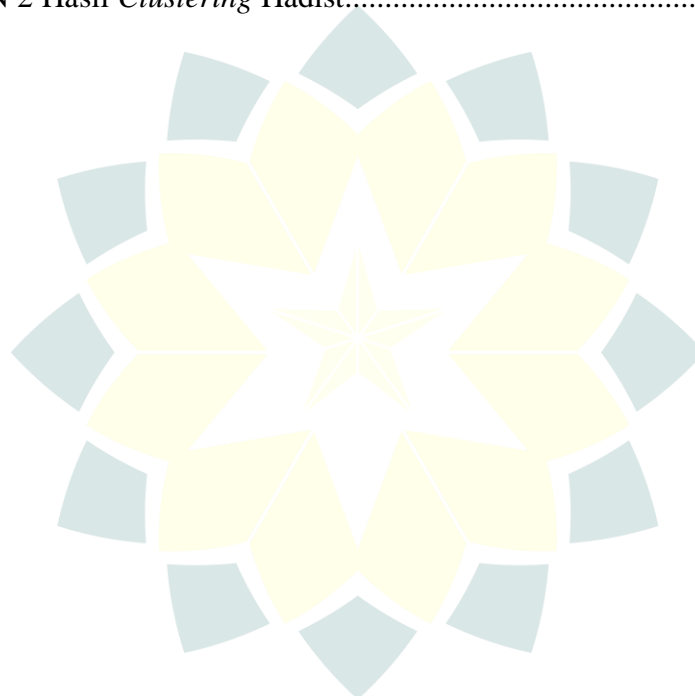
UNIVERSITAS ISLAM NEGERI  
SUNAN GUNUNG DJATI  
BANDUNG

## DAFTAR ISTILAH

- Auxiliary attributte* : Atribut tambahan dari sebuah data atau dokumen.
- Centroid* : Pusat dari sebuah kelompok atau *cluster*.
- Cluster* : Gugus.
- Clustering* : Proses penggugusan data tanpa label.
- Corpus* : Kumpulan kata yang terkandung dalam dokumen (kamus).
- Diskriminatif* : Bersifat diskriminasi (membeda-bedakan).
- Euclidean distance* : Metode perhitungan jarak.
- Frequency* : Ukuran jumlah dalam suatu peristiwa.
- Giny indeks* : Peluang keberadaan relatif sebuah dokumen
- IDF* : Nilai *invers* dari kemunculan sebuah dokumen.
- Matan :Teks hadist, berupa ucapan, perbuatan dan takrir yang terletak setelah sanad terakhir.
- Metadata* : Informasi yang mendeskripsikan sebuah dokumen.
- Sanad : Jalan yang menyampaikan kepada matan hadist.
- Side information* : Informasi yang dimiliki oleh sebuah dokumen.
- Stemming* : Proses pengembalian sebuah kata kedalam kata dasar.

## DAFTAR LAMPIRAN

LAMPIRAN 1 <i>Pseudocode</i> .....	47
LAMPIRAN 2 Hasil <i>Clustering</i> Hadist.....	53



uin

UNIVERSITAS ISLAM NEGERI  
SUNAN GUNUNG DJATI  
BANDUNG