

DAFTAR ISI

HALAMAN JUDUL

HALAMAN PERSETUJUAN

HALAMAN PENGESAHAN

PERNYATAAN KEASLIAN SKRIPSI

ABSTRAK

ABSTRACT

KATA PENGANTAR..... i

DAFTAR ISI..... iii

DAFTAR GAMBAR..... vii

DAFTAR TABEL ix

DAFTAR SIMBOL xi

DAFTAR LAMPIRAN xii

BAB I PENDAHULUAN..... 1

1.1 Latar Belakang Masalah 1

1.2 Rumusan Masalah..... 7

1.3 Batasan Masalah 7

1.4 Tujuan dan Manfaat Penelitian 8

1.5 Metode Penelitian 9

1.6 Sistematika Penulisan 9

BAB II LANDASAN TEORI 11

2.1 Penggalian Data (*Data Mining*) 11

2.2 Penggalian Teks (*Text Mining*)..... 12

2.3 Pengelompokan Teks (*Text Clustering*)..... 12

2.4 Algoritma *Clustering* (*Clustering Algorithm*) 13

2.5	Algoritma <i>Clustering K-Means</i>	16
2.6	Algoritma <i>Clustering K-Means++</i>	17
2.7	Pra-Pemrosesan Teks (<i>Text Preprocessing</i>)	19
2.7.1	<i>Case Folding</i>	19
2.7.2	<i>Tokenizing</i>	20
2.7.3	<i>Stopword Removal</i>	20
2.7.4	<i>Stemming</i>	20
2.8	<i>Term Weighting</i>	21
2.8.1	Metode <i>tf-idf</i>	21
2.9	Reduksi Dimensi Fitur	23
2.9.1	<i>Feature Extraction</i>	23
2.10	<i>Proximity Measure</i>	26
2.10.1	<i>Cosine Distance</i>	27
2.10.2	<i>Euclidean Distance</i>	28
2.11	<i>Davies-Bouldin Index (DBI)</i>	28
2.12	<i>Silhouette Coefficient (SC)</i>	30
2.13	<i>Python</i>	34
BAB III ALGORITMA <i>CLUSTERING K-MEANS++</i> PADA DATA TEKS TERJEMAHAN HADIS		35
3.1	Kitab Hadis	37
3.2	Pengumpulan Data (<i>Data Collection</i>).....	37
3.3	Analisis <i>Text Preprocessing</i>	40
3.3.1	<i>Tokenizing</i>	43
3.3.2	<i>Case Folding</i>	44
3.3.3	<i>Stopword Removal</i>	46
3.3.4	<i>Stemming</i>	47

3.4	<i>Analisis Term Weighting</i>	49
3.4.1	<i>Analisis Metode tf-idf</i>	50
3.5	<i>Reduksi Fitur</i>	55
3.5.1	<i>Principal Component Analysis (PCA)</i>	56
3.6	<i>Algoritma Clustering K-Means++</i>	58
3.6.1	<i>Skenario ke-1</i>	58
3.6.2	<i>Skenario ke-2</i>	59
BAB IV ANALISIS HASIL IMPLEMENTASI DAN PERCOBAAN CLUSTERING ALGORITMA K-MEANS++ PADA DATA TERJEMAHAN HADIS		61
4.1	<i>Dataset</i>	61
4.2	<i>Analisis Hasil Clustering Skenario ke-1</i>	62
4.2.1	<i>Analisis Hasil Clustering K-Means++ Menggunakan Cosine Distance Tanpa Reduksi</i>	63
4.2.2	<i>Analisis Hasil Clustering K-Means++ Menggunakan Euclidean Distance Tanpa Reduksi</i>	71
4.2.3	<i>Perbandingan Algoritma Clustering K-Means++ Menggunakan Cosine Distance dan Euclidean Distance Skenario Pertama</i>	78
4.3	<i>Hasil Variasi Reduksi menggunakan PCA</i>	80
4.3.1	<i>Reduksi Dimensi PCA pada Algoritma Clustering K-Means++ menggunakan Cosine Distance</i>	80
4.3.2	<i>Reduksi Dimensi PCA pada Algoritma Clustering K-Means++ menggunakan Euclidean Distance</i>	82
4.3.3	<i>Analisis Perbandingan Hasil Reduksi Menggunakan PCA</i>	85
4.4	<i>Analisis Hasil Clustering Skenario ke-2</i>	88
4.4.1	<i>Analisis Hasil Clustering K-Means++ Menggunakan Cosine Distance Setelah Reduksi</i>	88
4.4.2	<i>Analisis Hasil Clustering K-Means++ Menggunakan Euclidean Distance Setelah Reduksi</i>	94

4.4.3 Perbandingan Algoritma <i>Clustering K-Means++</i> Menggunakan <i>Cosine Distance</i> dan <i>Euclidean Distance</i> Skenario Kedua.....	101
4.5 Analisis Perbandingan <i>Clustering</i> Algoritma <i>K-Means++</i>	103
BAB V PENUTUP	105
5.1 Kesimpulan	105
5.2 Saran	106
DAFTAR PUSTAKA	107
RIWAYAT HIDUP	111
LAMPIRAN	

