

DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN.....	i
PERNYATAAN KEASLIAN SKRIPSI.....	i
ABSTRAK.....	i
ABSTRACT.....	i
KATA PENGANTAR.....	i
DAFTAR ISI.....	iii
DAFTAR GAMBAR.....	vii
DAFTAR TABEL	xi
DAFTAR SIMBOL.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	6
1.4 Tujuan dan Manfaat Penelitian	6
1.5 Metode Penelitian.....	7
1.6 Sistematika Penulisan.....	7
BAB II LANDASAN TEORI.....	10
2.1 <i>Data Mining</i>	10
2.2 <i>Text Mining</i>	11
2.3 <i>Data Preprocessing</i>	11

2.3.1	<i>Term Weighting</i>	12
2.3.2	Reduksi Fitur	12
2.4	<i>Proximity Measure</i>	13
2.5	<i>Clustering</i>	15
2.6	<i>K-Means</i>	17
2.7	<i>Outlier</i>	20
2.8	Evaluasi <i>Cluster</i>	24
2.9	<i>Python</i>	25
BAB III PERBANDINGAN PERFORMA METODE DETEKSI <i>OUTLIER</i>		
<i>DISTANCE-BASED</i> DAN <i>CLUSTER-BASED</i> PADA ALGORITMA <i>K-</i>		
<i>MEANS</i> DENGAN DATA TEKS HADIS		27
3.1	Data Hadis	28
3.2	Pengambilan Data	29
3.3	Analisis Data <i>Preprocessing</i>	31
3.3.1	<i>Case Folding</i>	31
3.3.2	<i>Tokenization</i>	33
3.3.3	<i>Stopwords Removal</i>	34
3.3.4	<i>Stemming</i>	36
3.4	Analisis <i>Term Weighting</i>	40
3.4.1	<i>Term Frequency</i> (tf).....	41
3.4.2	<i>Document Frequency</i> (df)	43
3.4.3	<i>Inverse Document Frequency</i> (idf).....	44
3.4.4	Pembobotan <i>Term frequency – Inverse document frequency</i>	46
3.5	<i>Principal Component Analysis</i> (PCA)	48

3.6	<i>Analisis Proximity Measure</i>	50
3.6.1	<i>Euclidean Distance</i>	50
3.6.2	<i>Cosine similarity</i>	50
3.7	<i>Analisis cluster</i>	51
3.8	<i>Deteksi Outlier</i>	53
3.9	<i>Analisis Evaluasi cluster</i>	59

BAB IV ANALISIS HASIL IMPLEMENTASI DAN PERCOBAAN PERBANDINGAN PERFORMA METODE DETEKSI *OUTLIER DISTANCE-BASED* DAN *CLUSTER-BASED* PADA ALGORITMA *K-MEANS* DENGAN DATA TEKS HADIS..... 61

4.1	<i>Pengumpulan Data</i>	61
4.2	<i>Preprocessing</i>	62
4.3	<i>Menggunakan Data Hasil Reduksi Fitur PCA</i>	69
4.3.1	<i>Hasil K-Means Dengan Euclidean Distance</i>	70
4.3.2	<i>Analisis Hasil Deteksi Outlier Distance Based</i>	73
4.3.3	<i>Analisis Hasil Deteksi Outlier Cluster Based</i>	77
4.3.4	<i>Hasil K-Means Dengan Cosine similarity</i>	80
4.3.5	<i>Analisis Hasil Deteksi Outlier Distance Based</i>	83
4.3.6	<i>Analisis Hasil K-Means Cluster Based</i>	87
4.4	<i>Menggunakan Data Hasil Term Frequency – Inverse Document Frequency</i> 90	
4.4.1	<i>Hasil K-Means Dengan Euclidean Distance</i>	90
4.4.2	<i>Analisis Hasil Deteksi Outlier Distance Based</i>	93
4.4.3	<i>Analisis Hasil Deteksi Outlier Cluster Based</i>	98
4.4.4	<i>Hasil K-Means Dengan Cosine similarity</i>	101

4.4.5 Analisis Hasil Deteksi <i>Outlier Distance Based</i>	105
4.4.6 Analisis Hasil Deteksi <i>Outlier Cluster Based</i>	110
BAB V PENUTUP	114
5.1 Kesimpulan.....	114
5.2 Saran.....	115
DAFTAR PUSTAKA	116
LAMPIRAN	119
RIWAYAT HIDUP	133

