

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERSETUJUAN TESIS.....	iii
PERNYATAAN BEBAS PLAGIASI.....	iv
PRAKATA.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR.....	ix
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN.....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan.....	6
1.5 Manfaat.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Tinjauan Pustaka Metode Representasi Teks.....	7
2.2 Tinjauan Pustaka Algoritma <i>Clustering</i>	13
2.3 Tinjauan Pustaka Metriks Evaluasi <i>Clustering</i>	18
BAB III LANDASAN TEORI.....	23
3.1 <i>Natural Language Processing (NLP)</i>	23
3.1.1 <i>Supervised Learning</i>	23
3.1.2 <i>Unsupervised Learning</i>	24
3.2 Representasi Teks.....	24
3.2.1 <i>Bag of Words</i>	25
3.2.2 <i>Word Embeddings (FastText)</i>	26
3.2.3 <i>Bidirectional Encoder Representations from Transformers (BERT)</i>	26
3.3 Uniform Manifold Approximation and Projection.....	29
3.4 Algoritma Klustering.....	30
3.4.1 <i>K-Means</i>	31
3.4.2 <i>Density-Based Spatial Clustering of Applications with Noise (DBSCAN)</i>	32
3.4.3 <i>Spectral Clustering</i>	34
3.5 Metriks Evaluasi.....	36
3.4.1 <i>Silhouette Score</i>	36
3.4.2 <i>Davies-Bouldin Index</i>	38
3.4.3 <i>Calinski-Harabasz Index (Variance Ratio Criterion)</i>	40
3.6 <i>Dataset Menu Restoran</i>	42
BAB IV METODOLOGI PENELITIAN.....	44

4.1	Studi Literatur	44
4.2	Alat & Bahan	44
4.2.1	<i>Alat</i>	44
4.2.2	<i>Bahan</i>	45
4.3	Pengumpulan Data	46
4.4	Prosedur kerja	46
4.4.1	<i>Alur Penelitian</i>	47
4.4.2	<i>Proses Akuisisi Data dan Implementasi Metode</i>	49
4.4.3	<i>Skema Evaluasi</i>	54
4.4.4	<i>Metode Perangkingan</i>	55
BAB V IMPLEMENTASI.....		58
5.1	Pengumpulan Data	58
5.2	Pra-Pemrosesan Data	61
5.2.1	<i>Membaca Dataset</i>	61
5.2.2	<i>Lowercasing</i>	62
5.2.3	<i>Deteksi Bahasa</i>	63
5.2.4	<i>Menghapus Duplikasi</i>	65
5.2.5	<i>Standarisasi Bahasa</i>	66
5.2.6	<i>Data Cleaning</i>	67
5.2.7	<i>Tokenisasi</i>	68
5.2.8	<i>Stopword Removal</i>	69
5.2.9	<i>Lemmatization</i>	70
5.3	Implementasi Representasi Teks.....	72
5.3.1	<i>Implementasi Bag of Word</i>	72
5.3.2	<i>Implementasi FastText</i>	73
5.3.3	<i>Implementasi BERT</i>	79
5.4	Implementasi Algoritma Clustering.....	81
5.4.1	<i>Implementasi K-Means</i>	81
5.4.2	<i>Implementasi DBSCAN</i>	84
5.4.3	<i>Implementasi Spectral Clustering</i>	88
BAB VI HASIL PENELITIAN DAN PEMBAHASAN		92
6.1	Analisis Hasil Representasi Teks	92
6.2	Analisis Perbandingan Berdasarkan Metrik Evaluasi	95
6.2.1	<i>Silhouette Score</i>	95
6.2.2	<i>Davies-Bouldin Index</i>	103
6.2.3	<i>Calinski-Harabasz Index</i>	110
6.2.4	<i>Waktu Komputasi</i>	117
6.3	Analisis Perbandingan Secara Keseluruhan.....	118
6.4	Analisis Visualisasi Hasil Clustering	131
6.5	Analisis Tambahan.....	140
BAB VII KESIMPULAN DAN SARAN		148
7.1	Kesimpulan	148
7.2	Saran.....	149

DAFTAR PUSTAKA	150
LAMPIRAN.....	155