

DAFTAR ISI

HALAMAN PENGESAHAN	i
HALAMAN PERNYATAAN	ii
HALAMAN MOTTO DAN PERSEMBAHAN	iii
PRAKATA	iv
DAFTAR ISI	vi
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
INTISARI	xiii
ABSTRACT	xiv
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Metodologi Penulisan	5
1.7 Sistematika Penulisan	6
II TINJAUAN PUSTAKA	8
III LANDASAN TEORI	13
3.1 Film	13
3.1.1 Genre film	13
3.1.2 Kata kunci alur	14
3.2 <i>Knowledge Discovery in Database</i>	14
3.2.1 Prapemrosesan data	15
3.2.2 <i>Stemming</i>	15

3.2.3	<i>Data mining</i>	16
3.3	Klasifikasi	17
3.3.1	Klasifikasi <i>k-Nearest Neighbors</i>	18
3.4	Perhitungan Jarak	19
3.4.1	<i>Cosine distance</i>	19
3.4.2	<i>Euclidean distance</i>	20
3.5	Teorema Bayes	21
3.6	Klasifikasi multilabel	22
3.6.1	<i>Multilabel k-Nearest Neighbor</i>	22
3.7	Analisis Karakteristik Dataset	25
3.7.1	Analisis karakteristik dasar	25
3.7.2	Analisis tingkat ketidakseimbangan label	26
3.8	Ekstraksi Fitur	28
3.8.1	<i>Bag-of-Words</i>	29
3.8.2	<i>Term Frequency - Inverse Document Frequency</i>	29
3.8.3	<i>Term-Document Matrix</i>	30
3.9	<i>Cross Validation</i>	30
3.10	Evaluasi Kinerja	31
3.10.1	<i>Jaccard similarity coefficient</i>	32
3.10.2	<i>Hamming loss</i>	32
3.10.3	Presisi	33
3.10.4	<i>Recall</i>	33
3.10.5	<i>F-measure</i>	34
3.10.6	Contoh perhitungan <i>Macro/Micro-Averaging</i>	34
IV	ANALISIS DAN PERANCANGAN	36
4.1	Analisis Permasalahan	36
4.2	Rancangan Umum Sistem	36
4.3	Data	38
4.4	Perancangan Prapemrosesan Data	39
4.5	Perancangan Pembentukan <i>Term-Document Matrix</i>	40
4.6	Perancangan Analisis Karakteristik Dataset	42
4.7	Klasifikasi <i>Multilabel k-Nearest Neighbor</i>	43
4.7.1	Perancangan pelatihan ML-KNN	43
4.7.2	Perancangan perhitungan probabilitas prior	44

4.7.3	Perancangan perhitungan probabilitas posterior	44
4.7.4	Perancangan pengujian ML-KNN	45
4.8	Perancangan <i>5-fold Cross Validation</i>	47
4.9	Perancangan Evaluasi Kinerja	48
V	IMPLEMENTASI	49
5.1	Lingkungan Implementasi	49
5.2	Implementasi Prapemrosesan Data	49
5.3	Implementasi Pembentukan <i>Term-Document Matrix</i>	50
5.4	Implementasi Analisis Karakteristik Dataset	51
5.4.1	Implementasi kardinalitas dan densitas label	51
5.4.2	Implementasi <i>Imbalance Ratio per Label</i>	51
5.4.3	Implementasi <i>Mean Imbalance Ratio</i>	52
5.4.4	Implementasi <i>Coefficient of Variation of IRLbl</i>	53
5.5	Implementasi Klasifikasi <i>Multilabel k-Nearest Neighbor</i>	53
5.5.1	Implementasi pelatihan ML-KNN	54
5.5.2	Implementasi perhitungan probabilitas prior	55
5.5.3	Implementasi perhitungan probabilitas posterior	55
5.5.4	Implementasi pengujian ML-KNN	57
5.6	Implementasi <i>5-fold Cross Validation</i>	58
5.7	Implementasi Evaluasi Kinerja	59
VI	HASIL DAN PEMBAHASAN	61
6.1	Hasil Prapemrosesan Data	61
6.2	Hasil Pembentukan <i>Term-Document Matrix</i>	62
6.3	Hasil Analisis Karakteristik Dataset	63
6.4	Hasil Klasifikasi <i>Multilabel k-Nearest Neighbor</i>	65
6.4.1	Hasil pelatihan ML-KNN	65
6.4.2	Hasil perhitungan probabilitas prior	65
6.4.3	Hasil perhitungan probabilitas posterior	66
6.4.4	Hasil pengujian ML-KNN	67
6.5	Hasil Evaluasi Kinerja	70
6.6	Pengaruh Nilai <i>k</i> terhadap Kinerja Klasifikasi	74
6.7	Pengaruh Nilai <i>min_df</i> terhadap Kinerja Klasifikasi	76

VII KESIMPULAN DAN SARAN	78
7.1 Kesimpulan	78
7.2 Saran	78
DAFTAR PUSTAKA	79
LAMPIRAN	82