

Daftar Isi

Lembar Pernyataan.....	ii
Lembar Pengesahan	iii
Abstrak	iv
Abstract	v
Lembar Persembahan	ii
Kata Pengantar	iii
Daftar Isi.....	iv
Daftar Gambar.....	vii
Daftar Tabel	viii
Daftar Istillah	ix
1 Bab I Pendahuluan.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	1
1.3 Tujuan.....	2
1.4 Batasan Masalah.....	2
1.5 Metodologi Penyelesaian Masalah	2
2 Bab II Tinjauan Pustaka.....	4
2.1 Text Mining	4
2.2 Text Categorization	4
2.3 Text Preprocessing	4
2.3.1 Case Folding	4
2.3.2 Tokenization.....	4
2.3.3 Stopword Removal.....	4
2.3.4 Stemming	5
2.4 Multinomial Naïve Bayes.....	7

2.5	Performansi Classifier	8
2.6	Document Oriented Database.....	9
2.7	MongoDB	9
2.8	GridFS	9
2.9	Replica Set.....	10
2.10	Sharding.....	11
2.10.1	Komponen Sharding	11
2.10.2	Shard Key.....	13
2.10.3	Splitting dan Balancing	13
2.11	Performansi Basisdata	14
2.11.1	Response time	14
2.11.2	Throughput.....	14
3	Bab III Perancangan dan Implementasi.....	15
3.1	Gambaran Umum Sistem	15
3.2	Analisis Data	16
3.3	Perancangan Sistem Kategorisasi.....	18
3.3.1	Preprocessing	18
3.3.2	Multinomial Naïve Bayes	21
3.4	Perancangan Arsitektur Sharding.....	26
3.4.1	Arsitektur I.....	26
3.4.2	Arsitektur II.....	27
3.4.3	Arsitektur III.....	28
3.4.4	Arsitektur IV	29
3.5	Spesifikasi Data.....	29
3.6	Spesifikasi Kebutuhan Sistem.....	30
3.6.1	Spesifikasi perangkat lunak.....	30

3.6.2	Spesifikasi perangkat keras	30
4	BAB IV Pengujian dan Analisis	31
4.1	Tujuan dan Skenario Pengujian.....	31
4.1.1	Tujuan Pengujian	31
4.1.2	Skenario Pengujian.....	31
4.2	Analisis dan Hasil Pengujian.....	32
4.2.1	Analisis Pengaruh Jumlah Data Latih terhadap Performansi.....	32
4.2.2	Analisis Pengaruh Jumlah <i>Shard</i> terhadap Performansi	36
5	Kesimpulan dan Saran	46
5.1	Kesimpulan.....	46
5.2	Saran	46
	Daftar Pustaka	47
	Lampiran A (Pengujian Sistem Kategorisasi).....	48
	Lampiran B (Pengujian Basisdata Dokumen).....	51