

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PENGESAHAN	ii
PERNYATAAN BEBAS PLAGIASI	iv
PRAKATA	v
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
INTISARI	xv
ABSTRACT	xvi
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Batasan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Tinjauan Pustaka	3
1.5 Metode Penelitian	4
1.6 Sistematika Penulisan	5
II DASAR TEORI	6
2.1 Skripsi	6
2.1.1 Abstrak	6
2.2 Data Mining	6
2.3 Text Mining	7
2.4 Machine Learning	8
2.5 Pengumpulan Data dari Web	9
2.5.1 Web Crawling	9
2.5.2 Web Scraping	9

2.6	Pra-pemrosesan Data Teks	10
2.6.1	Case Folding	10
2.6.2	Cleaning Text	11
2.6.3	Tokenize	11
2.6.4	Stopword Removal	12
2.6.5	Stemming	12
2.7	Visualisasi Data	12
2.7.1	Count Plot	13
2.7.2	Histogram	13
2.7.3	Box Plot	14
2.7.4	Line Plot	14
2.7.5	Word Cloud	14
2.8	Variabel Laten	15
2.9	Variabel Random	15
2.10	Probabilitas Bersyarat	16
2.11	Distribusi Probabilitas	17
2.11.1	Distribusi Gamma	18
2.11.2	Distribusi Beta	19
2.11.3	Distribusi Binomial	20
2.11.4	Distribusi Multinomial	20
2.11.5	Distribusi Dirichlet	21
2.12	Matriks	21
2.12.1	Jenis Matriks	22
2.12.2	Operasi Matriks	23
2.13	Hyperparameter Tuning	25
2.14	Pengelompokan Hierarki	26

III METODE PEMODELAN TOPIK MENGGUNAKAN LDA DAN NMF

	DENGAN PEMBOBOTAN KATA BOW SERTA TF-IDF	28
3.1	Pembobotan Kata	28
3.2	Bag of Words (BoW)	29
3.3	Term Frequency-Inverse Document Frequency (TF-IDF)	30
3.4	Pemodelan Topik	31
3.5	Latent Dirichlet Allocation (LDA)	32
3.5.1	Variational Bayes	35

3.5.2	Hyperparameter Model	37
3.6	Non-negative Matrix Factorization (NMF)	38
3.6.1	Projected Gradient Descent	39
3.6.2	Hyperparameter Model	40
3.7	Evaluasi Model	41
3.7.1	Coherence	41
3.7.2	Interpretabilitas	46
3.8	Alur Penelitian	47
IV	ANALISIS ABSTRAK SKRIPSI	49
4.1	Alat Analisis	49
4.2	Deskripsi Data	49
4.3	Proses Pengambilan Data	50
4.3.1	Web Crawling	50
4.3.2	Web Scraping	51
4.4	Pra-pemrosesan Data Teks	52
4.4.1	Case Folding	52
4.4.2	Cleaning Text	53
4.4.3	Tokenize	54
4.4.4	Stopword Removal	56
4.4.5	Stemming	57
4.5	Eksplorasi Data	59
4.5.1	Jumlah Skripsi Tiap Tahun	59
4.5.2	Jumlah Skripsi Tiap Pembimbing	60
4.5.3	Kata Kunci Skripsi	61
4.5.4	Jumlah Kata Abstrak Skripsi	61
4.5.5	Frekuensi Tiap Kata pada Abstrak Skripsi	62
4.6	Pembobotan Kata	63
4.6.1	Bag of Words (BoW)	64
4.6.2	Term Frequency-Inverse Document Frequency (TF-IDF)	65
4.7	Pemodelan Topik	65
4.7.1	Latent Dirichlet Allocation (LDA)	66
4.7.2	Non-negative Matrix Factorization (NMF)	71
4.8	Evaluasi Model	76
4.9	Interpretasi Topik	79

4.9.1	Topik 1: Analisis Valuasi Obligasi	80
4.9.2	Topik 2: Estimasi Value at Risk dan Penentuan Harga Opsi . .	83
4.9.3	Topik 3: Metode Statistika Klasik	85
4.9.4	Topik 4: Estimasi Cadangan Klaim dan Perhitungan Premi . .	87
4.9.5	Topik 5: Metode Pengendalian Kualitas	90
4.9.6	Topik 6: Variasi Model Regresi	92
4.9.7	Topik 7: Optimisasi Portofolio	95
4.9.8	Topik 8: Statistical Machine Learning	97
4.10	Eksplorasi Topik	99
4.10.1	Pengelompokan Hierarki Topik	100
4.10.2	Jumlah Skripsi Tiap Topik dan Kelompok	102
4.10.3	Tren Topik dan Kelompok Tiap Tahun	103
4.10.4	Proporsi Topik dan Kelompok Tiap Pembimbing	104
V	PENUTUP	108
5.1	Kesimpulan	108
5.2	Saran	109
	DAFTAR PUSTAKA	110
	LAMPIRAN	114
1	Contoh Data Skripsi Mahasiswa Program Studi S1 Statistika FMIPA UGM	114
2	Hasil Topik dan Kelompok Skripsi Mahasiswa Program Studi S1 Statistika FMIPA UGM	116
3	Code Script untuk Web Crawling dan Web Scraping	118
4	Code Script untuk Analisis Abstrak Skripsi	126