

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN TUGAS AKHIR.....	ii
HALAMAN PENGESAHAN PENGUJI SIDANG	iii
HALAMAN PERNYATAAN KEASLIAN	iv
KATA PENGANTAR	v
ABSTRAK	vi
DAFTAR ISI	vii
DAFTAR RUMUS	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
DAFTAR SINTAK	xvii

BAB 1 PENDAHULUAN

1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian.....	4
1.5. Metode Pendekatan.....	5
1.6. Sistematika Penulisan.....	6

BAB 2 LANDASAN TEORI

2.1. Data Mining	7
2.2. Text Mining	8
2.2.1. Case Folding	10
2.2.2. Tokenizing	10
2.2.3. Filtering / Stopword	10
2.2.4. Stemming	11
2.2.5. Tagging Dan Feature Selection	16

2.2.5.1. Feature Selection Data Mining	16
2.2.5.2. Feature Selection Text Mining	18
2.2.5.3. Chi Squared (χ^2)	21
2.2.5.3.1. Chi Squared Satu Sampel (χ^2)	21
2.2.5.3.2. Chi Squared Dua Sampel (χ^2)	22
2.2.5.3.3. Chi Squared k Sampel (χ^2)	23
2.2.5.3.4. Chi Squared Feature Selection	23
2.3. Klasifikasi	36
2.4. Microsoft Visual Basic	49
2.5. Microsoft SQL Server.	50
2.6. Model Perancangan Dan Pengujian	50
2.6.1. Model Perancangan	50
2.6.2. Pengujian	52

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

3.1. Bahasa Pemrograman yang Digunakan	53
3.2. Gambaran Kerja Sistem	53
3.2.1. Pembacaan Dokumen Teks	53
3.2.2. Pembuatan Classifier	54
3.2.2.1. Tokenizing Dan Case Folding	55
3.2.2.2. Stopword	56
3.2.2.3. Stemming	56
3.2.2.4. Feature Selection	58
3.2.2.5. Deklarasi Naïve Bayes	62
3.2.3. Klasifikasi Dokumen	63
3.2.3.1. Persentase Kata	64
3.2.2.5. Pengkategorian Dokumen	65
3.3. Interface	68
3.3.1. Form Utama	69

3.3.2. Form Persiapan Dokumen	71
3.3.3. Form Machine Learning	72
3.3.4. Form View Table	73
3.3.5. Form Chart	73
3.4 Database	74
3.4.1. Table Kamus	74
3.4.2. Table Imbuhan	74
3.4.3. Table Dokumen	75
3.4.4. Table Stopword	76
3.4.5. Table Stemming	77
3.4.1. Table Feature Selection	79
3.4.1. Table Hasil	82
 BAB 4 IMPLEMENTASI DAN PENGUJIAN	
4.1. Implementasi Sistem	83
4.1.1. Implementasi Sistem	83
4.1.2. Pembuatan Machine Learning	84
4.1.2.1. Tahap Tokenisasi Dan Case Folding	84
4.1.2.2. Tahap Stopword	85
4.1.2.3. Tahap Stemming	86
4.1.2.4. Feature Selection	89
4.1.2.5. Naïve Bayes Classifier	89
4.1.2.6. Klasifikasi Dokumen	90
4.1.2.6.1. Proses Presentase	91
4.1.2.6.2. Pengkategorian Dokumen	94
4.2. Pengujian Dan Analisis Sistem	97
4.2.1. Pengujian Pertama	98
4.2.2. Pengujian Kedua	99
4.2.3. Pengujian Ketiga	100

4.2.4. Pengujian Keempat	101
4.2.5. Pengujian Kelima	102
4.2.6. Pengujian Keenam	103
4.2.7. Pengujian Ketujuh	104
4.2.8. Pengujian Kedelapan	105
4.2.9. Pengujian Kesembilan	106
4.2.10. Pengujian Kesepuluh	107
4.2.11. Pengujian Rata-Rata	108
4.2.12. Rata-Rata Dan Hasil Pengujian	108

BAB 5 KESIMPULAN DAN SARAN

5.1. Kesimpulan	109
5.2. Saran	110

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR RIWAYAT HIDUP

DAFTAR RUMUS

Rumus 2.1 Rumus Information Gain	19
Rumus 2.2 Rumus Mutual Information	19
Rumus 2.3 Rumus Term Strength	20
Rumus 2.4 Rumus Chi Squared	20
Rumus 2.5 Rumus Chi Squared Satu Sampel	22
Rumus 2.6 Rumus Chi Squared Dua Sampel	22
Rumus 2.7 Rumus Chi Squared k Sampel	23
Rumus 2.8 Rumus Chi Squared Untuk Feature Selection	23
Rumus 2.9 Pengembangan Rumus Chi Squared Untuk Feature Selection	23
Rumus 2.10 Rumus Chi Squared $P(c_j)$	25
Rumus 2.11 Rumus Chi Squared $P(w_i)$	25
Rumus 2.12 Rumus Chi Squared $P(\bar{c}_j)$	26
Rumus 2.13 Rumus Chi Squared $\overline{ docs_j }$	26
Rumus 2.14 Rumus Chi Squared Penyederhanaan $P(\bar{c}_j)$	27
Rumus 2.15 Rumus Chi Squared $P(\bar{w}_i)$	27
Rumus 2.16 Rumus Chi Squared $\bar{n}_w$	27
Rumus 2.17 Rumus Chi Squared Penyederhanaan $P(\bar{w})$	28
Rumus 2.18 Rumus Chi Squared $P(w_i, c_j)$	28
Rumus 2.19 Rumus Chi Squared $P(\bar{w}_i, c_j)$	29
Rumus 2.20 Rumus Chi Squared $ docs_{\bar{i}j} $	29

Rumus 2.21 Rumus Chi Squared Penyederhanaan $P(\bar{w}_i, c_j)$	30
Rumus 2.22 Rumus Chi Squared $P(w_i, \bar{c}_j)$	30
Rumus 2.23 Rumus Chi Squared $P(\bar{w}_i, \bar{c}_j)$	30
Rumus 2.24 Rumus Chi Squared $ \overline{docs_{ij}} $	31
Rumus 2.25 Rumus Chi Squared Penyederhanaan $P(\bar{w}_i, \bar{c}_j)$	32
Rumus 2.26 Rumus Naïve Bayes Classifier V_{MAP}	39
Rumus 2.27 Rumus Naïve Bayes Classifier Penyederhanaan Ke-1 V_{MAP}	40
Rumus 2.22 Rumus Naïve Bayes Classifier Penyederhanaan Ke-2 V_{MAP}	40
Rumus 2.28 Rumus Naïve Bayes Classifier Penyederhanaan Ke-3 V_{MAP}	41
Rumus 2.29 Rumus Naïve Bayes Classifier Penyederhanaan Ke-4 V_{MAP}	41
Rumus 2.30 Rumus Naïve Bayes Classifier $P(v_j)$	42
Rumus 2.31 Rumus Naïve Bayes Classifier $P(w_i v_j)$	42

DAFTAR TABEL

Tabel 2.1 Aturan Untuk Partikel	12
Tabel 2.2 Aturan Untuk Kata Ganti Kepunyaan	13
Tabel 2.3 Aturan Awalan Pertama Orde Derivatif	13
Tabel 2.4 Awalan Orde Kedua Derivatif	14
Tabel 2.5 Aturan Derivatif Sufiks	14
Tabel 2.6 Hasil Uji Coba 5 Metode Feature Selection	21
Tabel 2.7 Kontingensi	22
Tabel 2.8 Contoh Dokumen Hasil Proses Stemming	32
Tabel 2.9 Contoh Hasil Perhitungan Chi Squared	36
Tabel 2.10 Praproses Dokumen Proses Naïve Bayes Classifier	44
Tabel 2.11 Contoh Perhitungan Naïve Bayes Classifier	45
Tabel 2.12 Contoh Praproses Dokumen Input	47
Tabel 2.13 Contoh Proses Pengkategorian Naïve Bayes Classifier	48
Tabel 3.1 Tabel Contoh Perhitungan Chi Squared	62
Tabel 3.2 Contoh Perhitungan Naïve Bayes Classifier	63
Tabel 3.3 Definisi Table mKamus	74
Tabel 3.4 Definisi Table mImbuhan	75
Tabel 3.5 Definisi Table mDokTraining	75
Tabel 3.6 Definisi Table mDokInput	76
Tabel 3.7 Definisi Table mStopword	76
Tabel 3.8 Definisi Table tempStopword	77
Tabel 3.9 Definisi Table tempStopwordInput	77
Tabel 3.10 Definisi Table tempStemming	78
Tabel 3.11 Definisi Table StemmingKataDasar	78
Tabel 3.12 Definisi Table tempStemmingInput	78
Tabel 3.13 Definisi Table StemmingInput	78
Tabel 3.14 Definisi Table ChiSquaredKataDasar	80

Tabel 3.15 Definisi Table NBKataDasar	81
Tabel 3.16 Definisi Table NBChiSquared	81
Tabel 3.17 Tabel Definisi Table tampungkata	81
Tabel 3.18 Tabel Definisi Table tempHasil	82

DAFTAR GAMBAR

Gambar 2.1 Flowchart Proses Kategorisasi Teks	9
Gambar 2.2 Contoh Proses Tokenizing	10
Gambar 2.3 Contoh Proses Stopword	11
Gambar 2.4 Contoh Proses Stemming	11
Gambar 2.5 Proses Naïve Bayes Classifier	38
Gambar 2.6 Model Perancangan Sistem Sekuensial Linier atau Waterfall	51
Gambar 3.1 Gambar Flowchart Proses Pembuatan Classifier	54
Gambar 3.2 Gambar Flowchart Proses Tokenizing	55
Gambar 3.3 Gambar Flowchart Proses Stemming	57
Gambar 3.4 Gambar Flowchart Proses Pembentukan Variable Hitung Chi Squared Dan Perhitungan Chi Squared	60
Gambar 3.5 Gambar Lanjutan Flowchart Proses Pembentukan Variable Hitung Chi Squared Dan Perhitungan Chi Squared	61
Gambar 3.6 Gambar Flowchart Proses Klasifikasi Dokumen	64
Gambar 3.7 Gambar Rancangan Tampilan Form Utama	69
Gambar 3.8 Gambar Rancangan Tampilan Persiapan Dokumen	71
Gambar 3.9 Gambar Rancangan Tampilan Machine Learning	72
Gambar 3.10 Gambar Rancangan Tampilan Table View	73
Gambar 3.11 Gambar Rancangan Tampilan Chart	73
Gambar 4.1 Gambar Form Persiapan Dokumen	83
Gambar 4.2 Gambar Form Machine Learning	84
Gambar 4.3 Gambar Form Klasifikasi Dokumen	91
Gambar 4.4 Grafik Pada Pengujian Pertama	98
Gambar 4.5 Grafik Pada Pengujian Kedua	99
Gambar 4.6 Grafik Pada Pengujian Ketiga	100
Gambar 4.7 Grafik Pada Pengujian Keempat	101
Gambar 4.8 Grafik Pada Pengujian Kelima	102

Gambar 4.9 Grafik Pada Pengujian Keenam	103
Gambar 4.10 Grafik Pada Pengujian Ketujuh	104
Gambar 4.11 Grafik Pada Pengujian Kedelapan	105
Gambar 4.12 Grafik Pada Pengujian Kesembilan	106
Gambar 4.13 Grafik Pada Pengujian Kesepuluh	107
Gambar 4.14 Grafik Pada Pengujian Rata-Rata	108

DAFTAR SINTAK

Sintak 4.1 Sintak Proses Pembuangan Karakter	86
Sintak 4.2 Sintak Split Pada Proses Tokenizing	85
Sintak 4.3 Query Pada Proses Stopword	86
Sintak 4.4 Sintak Pada Proses Stemming	87
Sintak 4.5 Sintak Pada Proses Stemming (Sambungan)	88
Sintak 4.6 Query Pada Proses Feature Selection Dengan Chi Squared	89
Sintak 4.7 Query Pada Proses Pembuatan Clasifier Dengan Naïve Bayes Classifier	90
Sintak 4.8 Query Pada Proses Persentase Feature	92
Sintak 4.9 Query Pada Proses Pengambilan Kata Feature	93
Sintak 4.10 Query Pada Proses Pengambilan Kata-Kata Naïve Bayes Berdasarkan Chi Squared	94
Sintak 4.11 Query Pada Proses Stopword Dokumen Input	95
Sintak 4.12 Query Pada Proses Pembobotan Kata Dokumen Input	96
Sintak 4.13 Sintak Pada Proses Penentuan Kategori Dokumen Input	97