

APLIKASI KLASIFIKASI DOKUMEN BERDASARKAN BIDANG MINAT MENGGUNAKAN METODE NAIVE BAYES

**(STUDI KASUS: Abstrak Karya Ilmiah pada Portal Artikel Tugas
Akhir Universitas Trunojoyo madura)**

Makhsusotin
(11.04.111.00023)

Pembimbing 1 : Achmad Jauhari, S.T.,M.Kom.

Pembimbing 2 : Sigit Susanto Putro, S.Kom.,M.Kom

ABSTRAK

Saat ini bentuk dokumen digital telah mengalami perkembangan yang sangat pesat, namun kumpulan dokumen yang banyak akan sulit untuk diketahui karakteristik informasinya. Banyak sekali macam-macam karya ilmiah pda portal artikel tugas akhir Universitas Trunojoyo Madura. Karya ilmiah tentang pendidikan, hukum, ekonomi, dan masih banyak sekali macam karya ilmiah yang lain. Dari berbagai macam karya ilmiah tersebut dapat dengan mudah sekali dilakukan pengkategorian dokumen secara manual oleh manusia. Akan tetapi jika dilakukan secara terkomputerisasi akan sangatlah sulit. Untuk itu penelitian ini menghasilkan sebuah aplikasi yang dapat mengkategorikan suatu dokumen. Dokumen yang digunakan berupa dokumen abstrak karya ilmiah penelitian berdasarkan bidang minat prodi Teknik Informatika yang berada di Universitas Trunojoyo Madura. Yaitu, bidang minat Sistem Informasi, Sistem Terdistribusi, Digital Multimedia dan CAI (*Computing Artificial Intelligence*). Dari penelitian ini dihasilkan kesimpulan bahwa dalam pembuatan Aplikasi Klasifikasi Dokumen Abstrak Karya Ilmiah ini telah berhasil mengklasifikasi dokumen sesuai dengan bidang minat yaitu, CAI(*Computing Artificial Intelligence*), Digital Multimedia, Sistem Terdistribusi dan Sistem Informasi dengan tingkat akurasi mencapai 77%.

Kata Kunci : *Klasifikasi, Preprocessing, Naive Bayes.*



KATA PENGANTAR



Puji syukur kehadiran Allah SWT karena berkat rahmat dan hidayah-Nya, penulis dapat menyelesaikan Tugas Akhir yang berjudul “Aplikasi Klasifikasi Dokumen Berdasarkan Bidang Minat Menggunakan Metode *Naive Bayes*”. Keberhasilan dalam menyelesaikan Tugas Akhir ini tidak terlepas dari bantuan, dukungan saran dan bimbingan dari berbagai pihak. Dan perkenankanlah pada kesempatan kali ini penulis menyampaikan rasa terimakasih sedalam-dalamnya kepada :

1. Allah SWT yang telah memberikan kami Petunjuk, Rahmat serta Hidayah-Nya dalam menyelesaikan pembuatan laporan Tugas Akhir ini. Shalawat serta salam semoga tetap tercurahkan kepada junjungan Nabi Besar Muhammad SAW.
2. Ibu dan Bapak saya tercinta. Bapak Suhri dan Ibu Hatirah. Adik Rois Jauhari dan semua keluarga besar saya tercinta. Terimakasih atas dorongan semangat dan doa dari kalian atas kelancaran dari awal proses pembuatan Tugas Akhir ini sampai selesai.
3. Bapak Jauhari selaku Kepala Prodi Teknik Infotmatika Fakultas Teknik Universitas Trunojoyo Madura.
4. Bapak Jauhari dan bapak Sigit selaku dosen pembimbing dimana dengan ketelatenan dan kesabarannya membimbing serta memberikan arahan sejak perencanaan tugas akhir hingga pada penulisan laporan tugas akhir ini.
5. Seluruh dosen Teknik Informatika yang telah memberikan ilmu-ilmunya selama perkuliahan. Semoga semua kebaikan mereka diganti dengan pahala dari-Nya.
6. Seluruh teman-teman Antik Eleven (Anak Teknik 2011). serta semua teman-teman yang lainnya yang tidak bisa penulis sebutkan satu-satu. Terimakasih atas dorongan semangat serta doa dari kalian. Semoga Allah membalas semua kebaikan kalian. Amin.



7. Teman-teman KKN 10. Terimakasih sudah pernah menjadi keluarga baru saya.
8. Terimakasih kepada sahabat-sahabat saya : Alifah, Owong, Veronica, papa Mbul, Rian dan juga uus. Terimakasih atas doa-doa kalian.
9. Seluruh keluarga kost Gang 6 : Bu wiwin, Pak Gani, mas Ardi, mas dwi, mbak Ely, Mia, Imah, Astri, mbok Riri, Mak Jizun, Ismi, Bertha, Friska, Aqnes, Rum, Zum, Sely, Utin, Sinta, Putri, Mega, dan Ayin . Terimakasih atas semangatnya untuk saya.
10. Terimakasih kepada Seluruh Geng Balap Sarjana : Ainun, herlina, Tsalist, Sany, mbak Nu dan Nur.

Dan terimakasih untuk pihak-pihak yang tidak bisa penulis sebutkan satu-satu. Terimakasih doa dan semangat yang telah diberikan. Penulis menyadari bahwa pengerjaan Tugas Akhir ini masih jauh dengan kesempurnaan. Oleh karena itu penulis menyatakan permohonan maaf atas segala kekurangan yang ada. Dengan segala kerendahan hati, berbagai saran dan kritik yang membangun akan sangat penulis harapkan dari rekan-rekan pembaca.

Bangkalan , 2016

Penulis



DAFTAR ISI

HALAMAN JUDUL	ii
LEMBAR PENGESAHAN	iii
HALAMAN PERNYATAAN ORISINALITAS.....	iv
HALAMAN MOTTO.....	v
LEMBAR PERSEMBAHAN	vi
ABSTRAK.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
DAFTAR ISTILAH.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	2
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Batasan Masalah.....	2
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan.....	4
BAB II STUDI PUSTAKA	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori.....	8
2.2.1 Definisi Klasifikasi	8
2.2.2 Pengertian <i>Text Mining</i>	8
2.2.3 Ruang Lingkup <i>Text Mining</i>	8
2.2.4 Ekstraksi Dokumen.....	11
2.2.5 <i>Term Frequency</i> dan <i>Inverse Document Frequency (TF-IDF)</i>	13
2.2.6 Metode <i>Naive Baiyes</i>	14
2.2.7 Model Probabilistic <i>Naive Bayes</i>	15
BAB III ANALISIS DAN PERANCANGAN SISTEM	23
3.1 Study Literatur	23
3.2 Pengumpulan Data	23
3.3 Perancangan Sistem.....	24
3.3.1 <i>Flowchart</i>	24



a. <i>Flowchart Sistem</i>	24
b. <i>Flowchart Preprocessing</i>	26
b. <i>Flowchart Uji Coba</i>	28
3.4 Perhitungan Manual.....	29
3.5 <i>Usecase Diagram</i>	33
3.6 Rancangan Tabel Sistem	34
BAB IV IMPLEMENTASI SISTEM	43
4.1 Implementasi	43
4.2 Lingkungan Uji Coba	49
4.3 Implementasi Program.....	50
4.4 Skenario Uji Coba Sistem.....	57
4.4.1 Uji Fungsionalitas	57
4.4.2 Uji Akurasi Sistem.....	61
4.5 Grafik Analisa Data	62
BAB V KESIMPULAN	67
5.1 Kesimpulan.....	67
5.2 Saran.....	67
DAFTAR PUSTAKA	69
DAFTAR RIWAYAT HIDUP.....	71

DAFTAR GAMBAR

Gambar 2.1 <i>Proses Tokenizing</i>	12
Gambar 2.2 <i>Process Filtering</i>	12
Gambar 2.2 <i>Process Stemming</i>	13
Gambar 3.1 <i>Flowchart Sistem</i>	25
Gambar 3.2 <i>Flowchart Preprocessing</i>	26
Gambar 3.3 <i>Flowchart Naive Bayes</i>	27
Gambar 3.4 <i>Flowchart Uji Coba</i>	28
Gambar 3.5 <i>Usecase Diagram Admin</i>	33
Gambar 3.6 <i>Usecase Diagram User</i>	34
Gambar 4.1 Data Hasil Upload Dokumen	44
Gambar 4.2 <i>Source code</i> Hasil Upload Dokumen	45
Gambar 4.3 <i>Source Code Preprocessing</i> Dokumen	48
Gambar 4.4 <i>Source Code</i> Hasil Klasifikasi	49
Gambar 4.5 Halaman Login Admin.....	50
Gambar 4.6 Hasil Upload Dokumen.....	51
Gambar 4.7 Hasil Data <i>Testing</i>	52
Gambar 4.8 Pengkategorian Manual.....	53
Gambar 4.9 Hasil <i>Preprocessing</i>	54
Gambar 4.10 Hasil Nilai <i>Probabilitas</i>	55
Gambar 4.11 Hasil Klasifikasi.....	56
Gambar 4.12 Grafik Data Abstrak angkatan 2010.....	63
Gambar 4.11 Grafik Data Abstrak angkatan 2011.....	64
Gambar 4.11 Grafik Data Abstrak angkatan 2012.....	64

DAFTAR TABEL

Tabel 2.1 Perbandingan Metode	6
Tabel 2.2 Dokumen Awal	18
Tabel 2.3 Kemunculan <i>Term</i>	19
Tabel 2.4 Pencarian Nilai Probabilitas.....	20
Tabel 3.1 Dokumen Awal	30
Tabel 3.2 Kemunculan <i>Term</i>	30
Tabel 3.3Pencarian Nilai Probabilitas.....	31
Tabel 3.4 Atribut Dokumen	34
Tabel 3.5 Hasil Klasifikasi.....	35
Tabel 3.6 Hasil <i>Token</i>	35
Tabel 3.7 Kamus Kata.....	36
Tabel 3.8 Prob_kamus.....	36
Tabel 3.9 Prob_kat_kamus.....	37
Tabel 3.10 tb_katadasar	37
Tabel 3.11 prob_kamus.....	35
Tabel 3.12 Prob_kat_kamus.....	38
Tabel 3.13 tb_katadasar	38
Tabel 3.14 tb_p_vi	38
Tabel 3.15 tb_stopword.....	39
Tabel 3.16 tb_testing.....	39
Tabel 3.17 tb_training.....	39
Tabel 3.18 text_pre	40
Tabel 3.19 text_pre_train.....	40
Tabel 3.20 Tabel <i>User</i>	41
Tabel 3.21 Tabel Evaluasi Klasifikasi.....	41
Tabel 4.1 Rencana Pengujian Sistem.	57
Tabel 4.2 Pengujian <i>Login</i>	58
Tabel 4.3 Pengujian Menu Utama.....	59
Tabel 4.4 Pengujian Manajemen <i>User</i>	59
Tabel 4.5 Pengujian Manajemen Admin.....	60
Tabel 4.6 Hasil Akhir.....	60
Tabel 4.7 Pengujian <i>Logout</i>	61
Tabel 4.8 Tabel Evaluasi Klasifikasi.....	61



DAFTAR ISTILAH

<i>Naive Bayes</i>	Metode yang digunakan dalam proses klasifikasi
<i>Tokenizing</i>	Pemisahan kata
<i>Filtering</i>	Penghapusan Kata Sambung
<i>Stemming</i>	Mencari kata dasar
<i>Probabilitas</i>	<i>Peluang Kemungkinan</i>