

“KLASIFIKASI BERITA OLAHRAGA BERBAHASA INDONESIA MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*”

Elmo Bahtiar Rahman

(14.04.1.1.1.00087)

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

Pembimbing II : Yoga Dwitya Pramudita, S.Kom., M.Cs.

ABSTRAK

Berita merupakan suatu informasi berdasarkan fakta atau laporan mengenai suatu kejadian yang sedang, atau telah terjadi. Berita dipublikasikan melalui media cetak, siaran, internet, maupun dari mulut ke mulut. kebutuhan informasi di Indonesia dalam bentuk digital baik berupa teks atau dokumen. Pengklasifikasian berita masih dilakukan secara manual sehingga rentan terjadi kesalahan. Banyaknya berita yang harus diklasifikasikan berdasarkan kategorinya dalam waktu yang singkat akan membuat editor mengalami kesulitan dan dapat menimbulkan kesalahan dalam pengkategorian dikarenakan editor harus membaca satu per satu berita-berita yang akan diklasifikasikan. Apabila pengklasifikasian berita dilakukan secara otomatis akan meminimalisir kesalahan pengkategorian dan lebih efisien. Dari implementasi sistem yang telah dilakukan, klasifikasi berita olahraga berbahasa Indonesia mampu diselesaikan dengan menggunakan metode *Support Vector Machine* dengan tingkat akurasi 95.72%. Berdasarkan nilai evaluasi *f1-score*, *Support Vector Machine* menghasilkan nilai evaluasi sebesar 95.72%, nilai *Precision* menghasilkan nilai evaluasi sebesar 95.72% dan nilai *Recall* menghasilkan nilai sebesar 95.72%. Hal tersebut menunjukkan bahwa dalam penelitian ini model *Support Vector Machine* menghasilkan ketepatan klasifikasi berita olahraga berbahasa Indonesia yang lebih baik.

Kata kunci: Klasifikasi, Berita Olahraga Berbahasa Indonesia, *Support Vector Machine*.

DAFTAR ISI

HALAMAN PENGESAHAN.....	v
HALAMAN PERNYATAAN ORISINALITAS	vii
HALAMAN PERSEMBAHAN.....	ix
ABSTRAK.....	xi
KATA PENGANTAR	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR	xvii
DAFTAR TABEL.....	xix
BAB I.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	2
1.5 Batasan Masalah	2
1.6 Metodologi Penelitian.....	3
1.7 Sistematika Penulisan	4
BAB II.....	5
2.1. Penelitian Sebelumnya.....	5
2.2. <i>Text Mining</i>	8
2.3. Klasifikasi	8
2.4. <i>Text Preprocessing</i>	8
2.5. <i>Algoritma Enhanced Confix Stripping Stemmer</i>	11
2.6. <i>Term Frequency</i>	16
2.7. <i>Support Vector Machine</i>	16
2.8. <i>Confusion Matrix</i>	27
BAB III	29
3.1. Analisa Sistem	29
3.2. Analisa Kebutuhan.....	29
3.3. Rancangan Sistem.....	30
3.4. Arsitektur Sistem	32
3.4.1. <i>Dataset</i> Berita Olahraga	32
3.4.2. <i>Preprocessing</i>	34
3.4.3. <i>Support Vector Machine</i>	35
BAB IV	37
4.1. Lingkungan Uji Coba.....	37

DAFTAR GAMBAR

Gambar 2.1 Contoh <i>Case Folding</i>	9
Gambar 2.2 Contoh <i>Tokenizing</i>	10
Gambar 2.3 Contoh <i>Filtering</i>	11
Gambar 2.4 Contoh <i>Stemming</i>	11
Gambar 2.5 Aturan Penghapusan Awalan untuk awalan “me-” ...	13
Gambar 2.6 Aturan Penghapusan Awalan untuk awalan “pe-”	13
Gambar 2.7 Aturan Penghapusan Awalan untuk awalan “be-”	14
Gambar 2.8 Aturan Penghapusan Awalan untuk awalan “te-”	15
Gambar 2.9 Aturan yang Dimodifikasi untuk gambar 2.5	15
Gambar 2.10 Aturan yang Dimodifikasi untuk gambar 2.6	15
Gambar 2.11 Mencari fungsi pemisah yang optimal untuk obyek yang bisa dipisahkan	17
Gambar 2.12 Memperbesar margin bisa meningkatkan probabilitas	18
Gambar 2.13 Data Spiral Yang Tidak Linier	22
Gambar 3.1 Arsitektur Klasifikasi Berita Olahraga Menggunakan <i>Support Vector Machine</i>	30
Gambar 3.2 <i>Dataset</i> Berita Olahraga	33
Gambar 3.3 Total Data Berita Olahraga	33
Gambar 3.4 <i>Flowchart Preprocessing</i>	34
Gambar 3.5 <i>Flowchart</i> Klasifikasi <i>Support Vector Machine</i>	35
Gambar 4.1 Hasil Jumlah Data <i>Training</i>	39
Gambar 4.2 Hasil Jumlah Data <i>Testing</i>	39
Gambar 4.3 Hasil <i>Case Folding</i>	40
Gambar 4.4 Hasil <i>Filtering</i> dan <i>Stemming</i>	43
Gambar 4.5 Hasil Klasifikasi <i>Support Vector Machine</i>	47
Gambar 4.6 Halaman Form Klasifikasi	48
Gambar 4.7 Halaman Data <i>Training</i>	49
Gambar 4.8 Halaman Detail Data <i>Training</i>	49
Gambar 4.9 Halaman Data <i>Testing</i>	50
Gambar 4.10 Halaman Detail Data <i>Testing</i>	50
Gambar 4.11 Contoh Data Hasil <i>Preprocessing</i> Data <i>Training</i>	52

DAFTAR TABEL

Tabel 2.1 Penelitian Sebelumnya	7
Tabel 2 2 Fungsi Kernel	24
Tabel 2 3 Operasi Logika AND	24
Tabel 2 4 Operasi Logika XOR	25
Tabel 2.5 Confusion Matriks.....	27
Tabel 4 1 Lingkungan Uji Coba Sistem	37
Tabel 4.2 Hasil Uji Coba Mamad Taufik.....	68
Tabel 4.3 Hasil Uji Coba Aditya Wahyu Pratama	80
Tabel 4 4 Rangkuman Hasil Uji Coba.....	90