

KLASIFIKASI BERITA OLAHRAGA MENGUNAKAN METODE NAÏVE BAYES DENGAN ENHANCED CONFIX STRIPPING STEMMER

Nurul Makhmud
(13.04.1.1.1.00092)

Pembimbing 1 : Sigit Susanto Putro, S.Kom., M.Kom.
Pembimbing 2 : Yoga Dwitya Pramudita, S.Kom., M.Cs.

ABSTRAK

Dokumen berita olahraga dalam bentuk *web* saat ini memiliki jumlah yang besar dalam kurun waktu yang singkat. Besarnya jumlah dokumen berita olahraga membuat pengelompokan berita kedalam beberapa kategori menjadi hal yang sangat penting untuk diterapkan. Hal tersebut dilakukan dengan tujuan agar berita olahraga tersusun sesuai dengan kategorinya masing-masing. Berita dapat dikelompokkan secara manual oleh manusia, akan tetapi hal tersebut membutuhkan waktu yang lama untuk melakukan kategorisasi. Klasifikasi merupakan salah satu cara agar dapat melakukan pemberian kategori secara otomatis pada suatu berita. Tujuan dilakukannya klasifikasi adalah untuk mempercepat dan mempermudah dalam pemberian kategori pada suatu dokumen berita olahraga, sehingga dapat meningkatkan efisiensi waktu. Pada penelitian ini metode klasifikasi yang akan digunakan adalah dengan menggunakan metode *Naïve Bayes Classifier* karena komputasinya yang sederhana dan cepat. Pengujian menggunakan 18 berita olahraga yang dipilih secara acak oleh *user* atau *tester*, dari 18 berita yang diujikan terdapat 14 berita yang bernilai benar atau relevan dengan analisa yang dilakukan *user/tester* pada berita uji. Dari penelitian ini dapat disimpulkan bahwa Aplikasi Klasifikasi Berita Olahraga menggunakan Metode *Naïve Bayes* dengan *Enhanced Confix Striping Stemmer* mampu untuk mengklasifikasi berita olahraga sesuai dengan kategori masing-masing, seperti Sepak Bola, Basket, Raket, Formula 1, Moto GP dan olahraga lainnya dengan keakuratan sebesar 77%.

Kata Kunci : *Klasifikasi, Berita Olahraga, Naïve Bayes Classifier*



Abstract

Web-based sports news currently has a considerable amount of documents. News documents need to be grouped into multiple categories for easy access. The goal is that sports news is structured according to the specified category. News can be grouped manually by humans, but it takes a long time to categorize if it involves large documents. Classification method is proposed in this research to categorize automatically news document. The purpose of doing the classification is to accelerate and simplify the granting of categories, thereby increasing the efficiency of time. In this research the classification method used is Naïve Bayes Classifier method. This method is considered to have a simple and fast computing. From the test using 18 sports news randomly selected by the user or tester, there are 14 news stories that are true or relevant to the analysis by the user or the tester on the test news. This study concludes that the Sports News Classification Application using the Naïve Bayes Method with Enhanced Confix Striping Stemmer is able to classify sports news according to their respective categories, such as Football, Basket, Racquet, Formula 1, Moto GP and other sports with accuracy of 77%.

Keywords: *Classification, Sports News, Naïve Bayes Classifier*



KATA PENGANTAR

Alhamdulillah puji syukur senantiasa penulis ucapkan atas nikmat, karunia, rahmat, serta hidayah Allah SWT, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul :

“ KLASIFIKASI BERITA OLAHRAGA MENGUNAKAN METODE NAÏVE BAYES DENGAN ENHANCED CONFIX STRIPPING STEMMER”

Dalam penyusunan Tugas Akhir ini, penulis menyadari bahwa semua ini tidak lepas dari peran dan dukungan dari berbagai pihak yang telah dengan tulus senantiasa memberikan dukungan, bimbingan, dan motivasi, hingga akhirnya Tugas Akhir ini dapat terselesaikan. Oleh sebab itu penulis tak lupa menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT yang telah memberikan petunjuk, rahmat, serta hidayah-Nya dalam menyelesaikan seluruh proses dalam Tugas Akhir ini. Shalawat serta salam semoga selalu tercurahkan kepada junjungan Nabi Besar Muhammad SAW yang menjadi teladan dan telah menyampaikan petunjuk-Nya.
2. Kedua orang tuaku, serta seluruh keluarga besar yang selalu memberikan dukungan dan semangat, baik dalam bentuk materi, moril, maupun doa-doa yang selalu dipanjatkan kepada Allah SWT. Semoga semuanya senantiasa sehat dan selalu dalam lindungan Allah SWT.
3. Bapak Sigit Susanto Putro, S.Kom., M.Kom selaku dosen pembimbing I dan Yoga Dwitya Pramudita, S.Kom., M.Cs. selaku dosen pembimbing II. Terimakasih karena bapak telah banyak meluangkan waktu untuk membimbing dan memberikan arahan kepada penulis, sehingga mampu terselesaikannya Tugas Akhir ini.
4. Bapak dan Ibu Dosen Fakultas Teknik, khususnya Prodi Teknik Informatika Universitas Trunojoyo Madura.
5. Teman-teman Teknik Informatika UTM dan Asisten Praktikum Lab Teknik Informatika atas bimbingan dan dukungannya selama ini.
6. Teman-teman yang senasib dan seperjuangan.



7. Dan untuk semua pihak yang tidak dapat disebutkan satu persatu yang turut berperan dalam Tugas Akhir ini.

Akhir kata, penulis mengucapkan terima kasih dan mohon maaf apabila terdapat kesalahan dalam pengetikan, penulisan, dan penyampaian dalam Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat dan dapat dikembangkan dengan lebih baik nantinya.

Bangkalan, Juli 2017

Penulis



DAFTAR ISI

	Hal.
COVER	iii
HALAMAN JUDUL	iv
HALAMAN PENGESAHAN	v
HALAMAN MOTTO	ix
HALAMAN PERSEMBAHAN	xi
ABSTRAK	xiii
KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xx
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah	2
1.3. Tujuan dan Manfaat	3
1.3.1. Tujuan	3
1.3.2. Manfaat	3
1.4. Batasan Masalah	3
1.5. Metodologi Penelitian	3
1.6. Sistematika Penulisan	4
BAB II STUDI PUSTAKA	7
2.1. Penelitian Sebelumnya	7
2.2. Landasan Teori	14
2.2.1. Text Mining	14
2.2.2. Klasifikasi	14
2.2.3. Text Preprocessing	14
2.2.4. Algoritma Stemming Nazief dan Adriani	17
2.2.5. Algoritma Confix Stripping Stemmer	21
2.2.6. Algoritma Enhanced Confix Stripping Stemmer	22
2.2.7. Term Frequency	24
2.2.8. Naïve Bayes Classifier	24
2.2.9. Confusion Matrix	26
BAB III PERANCANGAN SISTEM	27
3.1. Alur Program	27
3.2. Arsitektur Sistem	28
3.3. Kerangka Desain Interface Sistem	29
3.3.1. Interface Utama Aplikasi	29



3.3.2	Interface Hasil Klasifikasi.....	30
3.4.	Metode	31
3.4.1	Flowchart Naïve Bayes Classifier	31
3.4.2	Flowchart Preprocessing	32
BAB IV	IMPLEMENTASI SISTEM.....	37
4.1.	Spesifikasi Perangkat.....	37
4.2.	Implementasi User Interface dan Kode Program	37
4.2.1	Halaman Form Klasifikasi	37
4.2.2	Halaman Hasil Klasifikasi.....	39
4.2.3	Halaman Data Berita Latih.....	42
4.2.4	Halaman Tabel Berita Latih	42
4.2.5	Halaman Tabel Kategori	47
4.2.6	Halaman Tabel Token	48
4.3	Pengujian	49
BAB V	PENUTUP	61
5.1.	Kesimpulan	61
5.2.	Saran.....	61
DAFTAR PUSTAKA.....		63
TENTANG PENULIS		65



DAFTAR GAMBAR

	Hal.
Gambar 2. 1 Contoh Case Folding	15
Gambar 2. 2 Contoh Tokenizing	15
Gambar 2. 3 Contoh Filtering	16
Gambar 2. 4 Contoh Stemming	17
Gambar 2. 5 Tahapan Klasifikasi Dokumen menggunakan Naïve Bayes Classifier	26
Gambar 2. 6 Confusion Matriks	26
 Gambar 3. 1 Flowchart Klasifikasi Berita Olahraga	 27
Gambar 3. 2 Arsitektur Klasifikasi Berita Olahraga Menggunakan Naïve Bayes Classifier	28
Gambar 3. 3 Interface Utama Aplikasi.....	29
Gambar 3. 4 Interface Hasil Klasifikasi	30
Gambar 3. 5 Flowchart Naïve Bayes Classifier	31
Gambar 3. 6 Flowchart Preprocessing	32
Gambar 3. 7 Flowchart Enhanced Confix Stripping Stemmer	33
 Gambar 4. 1 Tampilan Form Klasifikasi.....	 38
Gambar 4. 2 Kode Program Upload File.....	39
Gambar 4. 3 Tampilan Peringatan Ekstensi	39
Gambar 4. 4 Tampilan Hasil Klasifikasi	40
Gambar 4. 5 Kode Program Perhitungan Naïve Bayes	42
Gambar 4. 6 Tampilan Data Berita Latih	42
Gambar 4. 7 Tampilan Data Berita Latih	43
Gambar 4. 8 Kode Program Case Folding	43
Gambar 4. 9 Kode Program Tokenizing	43
Gambar 4. 10 Kode Program Filtering	44
Gambar 4. 11 Kode Program Stemming	45
Gambar 4. 12 Kode Program Indexing	46
Gambar 4. 13 Kode Program Model Probabilitas	47
Gambar 4. 14 Tampilan Data Kategori	47
Gambar 4. 15 Tampilan Data Token.....	48
Gambar 4. 16 Daftar Berita Uji.....	56



DAFTAR TABEL

	Hal.
Tabel 2. 1 Penelitian Sebelumnya.....	11
Tabel 2. 2 Penelitian Sebelumnya (Lanjutan)	12
Tabel 2. 3 Penelitian Sebelumnya (Lanjutan)	13
Tabel 2. 4 Aturan Pemenggalan Awalan	18
Tabel 2. 5 Modifikasi dan Tambahan Aturan Oleh Confix Stripping Stemmer.....	21
Tabel 2. 6 Modifikasi dan Tambahan Aturan Oleh Enhanced Confix Stripping Stemmer	23
 Tabel 4. 1 Spesifikasi Perangkat pada laptop	 37
Tabel 4. 2 Form Hasil Pengujian	57

