

KLASIFIKASI TEKS DOKUMEN BERITA OLAHRAGA MENGUNAKAN METODE ALGORITMA *K-NEAREST NEIGHBOR (KNN)*

Gloria Theresia Febryanti

(160411100059)

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

Dosen Pembimbing 2 : Mula'ab, S.Si., M.Kom.

ABSTRAK

Berita merupakan suatu informasi penting bagi kehidupan manusia. Berita banyak disebarakan melalui media cetak seperti koran dan majalah dan juga melalui media online seperti sosial media dan portal berita. Berita yang sangat diminati dari berbagai kalangan maupun usia yaitu berita olahraga. Dengan banyaknya berita olahraga yang disajikan oleh penyedia situs maka dibutuhkan pencermatan lebih untuk mendapatkan informasi berita olahraga sesuai dengan yang diinginkan. Tujuan dari penelitian ini untuk mengetahui tingkat akurasi dalam mengklasifikasikan teks dokumen berita olahraga dengan menggunakan metode *K – Nearest Neighbor (KNN)*. Data yang digunakan pada penelitian ini yaitu berita olahraga yang diambil dari portal berita <https://www.tempo.co/>. Data yang digunakan adalah 350 data berita olahraga yang terdiri dari 7 kelas yaitu sepakbola, bulutangkis, voli, renang, atletik, basket, motogp dan pada masing-masing kelas menggunakan 50 data. Data berita olahraga dilabeli secara manual berdasarkan isi berita. Setelah itu data akan dilakukan tahap *preprocessing* yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*. Data yang telah dipreprocessing akan dilakukan seleksi fitur menggunakan *information gain* untuk mengetahui bobot fitur pada data. Kemudian diklasifikasikan menggunakan metode algoritma *KNN* dan dilanjutkan pengujian menggunakan *K-Folding* dengan $K=5$ serta *confusion matrix* untuk mengetahui tingkat *akurasi*, *presisi*, dan *recall*. Berdasarkan pengujian yang dilakukan menggunakan metode *K-Nearest Neighbor (KNN)* dengan seleksi fitur *Information Gain* diperoleh nilai akurasi terbaik pada saat nilai k -tetangga 7 dengan

nilai ambang batas IG sama dengan 0.5 dan K-Fold 1 yaitu sebesar 51.42%, presisi 61.11%, recall 51.42% .

Kata Kunci : *klasifikasi, text preprocessing, seleksi fitur, Information Gain, K-Nearest Neighbor, K-Fold Cross Validation*

DAFTAR ISI

SKRIPSI	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN ORISINALITAS	iii
ABSTRAK	iv
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	ixi
DAFTAR TABEL	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah	2
1.2.1 Masalah Penelitian.....	2
1.2.2 Metode Usulan.....	3
1.2.3 Pertanyaan Penelitian.....	3
1.3 Tujuan dan Manfaat Penelitian.....	3
1.3.1 Tujuan Penelitian.....	3
1.3.2 Manfaat Penelitian.....	3
1.4 Batasan Masalah	3
1.5 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Klasifikasi	5
2.2 Text Preprocessing.....	5
2.3 TF-IDF	5
2.4 K-Nearest Neighbor (KNN).....	6
2.5 Information Gain	6
2.6 Proses Mining.....	7
2.7 Evaluasi.....	8
2.8 Penelitian Sebelumnya.....	8
BAB III METODE USULAN	13
3.1 Metodologi Penelitian	13
3.2 Perancangan Sistem.....	15
3.2.1 Arsitektur Sistem	15

3.2.2 Flowchart Information Gain.....	18
3.2.3 Flowchart KNN	19
3.3 Dataset.....	20
3.4 Skenario Pengujian.....	21
3.4.1 Skenario 1	22
3.4.2 Skenario 2	23
3.4.3 Skenario 3	23
3.4.4 Skenario 4	24
3.4.5 Skenario 5	24
3.4.6 Skenario 6	25
3.4.7 Skenario 7	25
3.4.8 Skenario 8	26
3.4.9 Skenario 9	26
3.4.10 Skenario 10	27
3.4.11 Skenario 11	27
3.4.12 Skenario 12	28
3.4.13 Skenario 13	29
3.4.14 Skenario 14	29
3.4.15 Skenario 15	30
3.4.16 Skenario 16	30
3.4.17 Skenario 17	31
3.4.18 Skenario 18	31
3.4.19 Skenario 19	32
3.4.20 Skenario 20	32
3.4.21 Skenario 21	33
3.4.22 Skenario 22	33
3.4.23 Skenario 23	34
3.4.24 Skenario 24	34
3.4.25 Skenario 25	35
BAB IV IMPLEMENTASI SISTEM.....	36
4.1 Lingkungan Uji Coba	36
4.2 Pembuatan Sistem.....	36
4.3 Hasil Implementasi Sistem	41
4.4 Hasil Pengujian	45

4.4.1 Hasil Pengujian Skenario 1	45
4.4.2 Hasil Pengujian Skenario 2	46
4.4.3 Hasil Pengujian Skenario 3	46
4.4.4 Hasil Pengujian Skenario 4	47
4.4.5 Hasil Pengujian Skenario 5	48
4.4.6 Hasil Pengujian Skenario 6	48
4.4.7 Hasil Pengujian Skenario 7	49
4.4.8 Hasil Pengujian Skenario 8	49
4.4.9 Hasil Pengujian Skenario 9	50
4.4.10 Hasil Pengujian Skenario 10	51
4.4.11 Hasil Pengujian Skenario 11	51
4.4.12 Hasil Pengujian Skenario 12	52
4.4.13 Hasil Pengujian Skenario 13	52
4.4.14 Hasil Pengujian Skenario 14	53
4.4.15 Hasil Pengujian Skenario 15	54
4.4.16 Hasil Pengujian Skenario 16	54
4.4.17 Hasil Pengujian Skenario 17	55
4.4.18 Hasil Pengujian Skenario 18	55
4.4.19 Hasil Pengujian Skenario 19	56
4.4.20 Hasil Pengujian Skenario 20	57
4.4.21 Hasil Pengujian Skenario 21	57
4.4.22 Hasil Pengujian Skenario 22	58
4.4.23 Hasil Pengujian Skenario 23	59
4.4.24 Hasil Pengujian Skenario 24	59
4.4.25 Hasil Pengujian Skenario 25	60
4.5 Analisa Pengujian	60
BAB V PENUTUP	62
5.1 Kesimpulan	62
5.2 Saran	62
REFERENSI	63
LAMPIRAN 1 Kode Program	67

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi K-Fold Cross Validation = 5	7
Gambar 3.1 Tahapan Penelitian	13
Gambar 3.2 Arsitektur Sistem	15
Gambar 3.3 Flowchart Information Gain	18
Gambar 3.4 Flowchart KNN	19
Gambar 4.1 Tampilan awal sistem	42
Gambar 4.2 Tampilan isi berita dan hasil klasifikasi	42
Gambar 4.3 Tampilan Login	43
Gambar 4.4 Tampilan menu dashboard	43
Gambar 4.5 Tampilan menu data	44
Gambar 4.6 Tampilan menu metode	45

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	8
Tabel 2.2 Penelitian Sebelumnya	12
Tabel 3.1 Dataset	20
Tabel 3.2 Skenario 1.....	23
Tabel 3.3 Skenario 2.....	23
Tabel 3.4 Skenario 3.....	24
Tabel 3.5 Skenario 4.....	24
Tabel 3.6 Skenario 5.....	25
Tabel 3.7 Skenario 6.....	25
Tabel 3.8 Skenario 7.....	26
Tabel 3.9 Skenario 8.....	26
Tabel 3.10 Skenario 9.....	27
Tabel 3.11 Skenario 10.....	27
Tabel 3.12 Skenario 11.....	28
Tabel 3.13 Skenario 12.....	29
Tabel 3.14 Skenario 13.....	29
Tabel 3.15 Skenario 14.....	30
Tabel 3.16 Skenario 15.....	30
Tabel 3.17 Skenario 16.....	31
Tabel 3.18 Skenario 17.....	31
Tabel 3.19 Skenario 18.....	32
Tabel 3.20 Skenario 19.....	32
Tabel 3.21 Skenario 20.....	33
Tabel 3.22 Skenario 21.....	33
Tabel 3.23 Skenario 22.....	34
Tabel 3.24 Skenario 23.....	34
Tabel 3.25 Skenario 24.....	35
Tabel 3.26 Skenario 25.....	35
Tabel 4.1 Spesifikasi Perangkat Keras.....	36
Tabel 4.2 Hasil pengujian skenario 1	45
Tabel 4.3 Hasil pengujian skenario 2	46

Tabel 4.4 Hasil pengujian skenario 3	47
Tabel 4.5 Hasil pengujian skenario 4	47
Tabel 4.6 Hasil pengujian skenario 5	48
Tabel 4.7 Hasil pengujian skenario 6	49
Tabel 4.8 Hasil pengujian skenario 7	49
Tabel 4.9 Hasil pengujian skenario 8	50
Tabel 4.10 Hasil pengujian skenario 9	50
Tabel 4.11 Hasil pengujian skenario 10	51
Tabel 4.12 Hasil pengujian skenario 11	52
Tabel 4.13 Hasil pengujian skenario 12	52
Tabel 4.14 Hasil pengujian skenario 13	53
Tabel 4.15 Hasil pengujian skenario 14	53
Tabel 4.16 Hasil pengujian skenario 15	54
Tabel 4.17 Hasil pengujian skenario 16	55
Tabel 4.18 Hasil pengujian skenario 17	55
Tabel 4.19 Hasil pengujian skenario 18	56
Tabel 4.20 Hasil pengujian skenario 19	57
Tabel 4.21 Hasil pengujian skenario 20	57
Tabel 4.22 Hasil pengujian skenario 21	58
Tabel 4.23 Hasil pengujian skenario 22	58
Tabel 4.24 Hasil pengujian skenario 23	59
Tabel 4.25 Hasil pengujian skenario 24	60
Tabel 4.26 Hasil pengujian skenario 25	60

