



REPUBLIK INDONESIA
KEMENTERIAN HUKUM DAN HAK ASASI MANUSIA

SURAT PENCATATAN CIPTAAN

Dalam rangka perlindungan ciptaan di bidang ilmu pengetahuan, seni dan sastra berdasarkan Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta, dengan ini menerangkan:

Nomor dan tanggal permohonan : EC00201823297, 9 Juni 2018

Pencipta

Nama : **Dr. Okfalisa, ST, MS.c**
Alamat : Jl. H.R Soebrantas No. 115 Km. 15 Panam, Pekanbaru, Riau, 28293
Kewarganegaraan : Indonesia

Pemegang Hak Cipta

Nama : **Dr. Okfalisa, ST, MS.c**
Alamat : Jl. H.R Subrantas No. 115 Km. 15 Panam, Pekanbaru, Riau, 28293
Kewarganegaraan : Indonesia
Jenis Ciptaan : **Buku Pelajaran**
Judul Ciptaan : **TF-IDF Dan Naive Bayes Untuk Filtering Dan Monitoring Diskusi Online**

Tanggal dan tempat diumumkan untuk pertama kali di wilayah Indonesia atau di luar wilayah Indonesia : 25 Agustus 2016, di Pekanbaru

Jangka waktu perlindungan : Berlaku selama hidup Pencipta dan terus berlangsung selama 70 (tujuh puluh) tahun setelah Pencipta meninggal dunia, terhitung mulai tanggal 1 Januari tahun berikutnya.

Nomor pencatatan : 000113793

adalah benar berdasarkan keterangan yang diberikan oleh Pemohon.

Surat Pencatatan Hak Cipta atau produk Hak terkait ini sesuai dengan Pasal 72 Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta.

a.n. MENTERI HUKUM DAN HAK ASASI MANUSIA
DIREKTUR JENDERAL KEKAYAAN INTELEKTUAL

Dr. Freddy Harris, S.H., LL.M., ACCS.
NIP. 196611181994031001



Testimonial:

- *"Buku ini sangat tepat untuk Mahasiswa khususnya dibidang Pendidikan Komputer dan Teknik Informatika dalam mempelajari dan mengefektivaskan konsep dasar e-learning melalui diskusi on-line. Penerapan Metode TF-IDF dan Naive Bayes menambah optimalitas pelaksanaan diskusi tersebut. Selain mahasiswa, buku ini juga dapat dijadikan pedoman bagi instruktur ataupun tenaga didik dalam memonitor dan mengevaluasi pelaksanaan diskusi on-line dalam proses pembelajaran".*

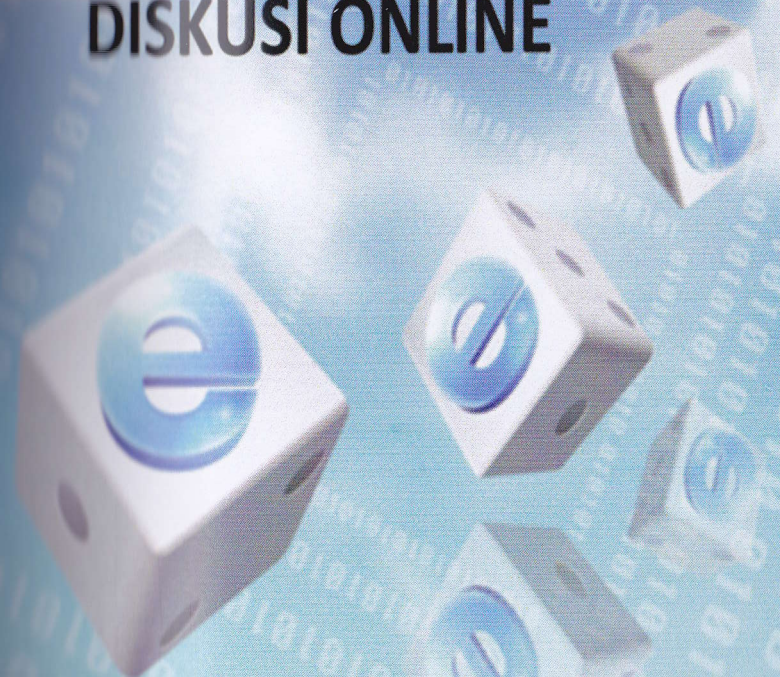
Dr. Saktioto, Mphil (Dosen FMIPA UNRI)

Penerbit: Al-Mujtahadah Press
Jl. Handayani Gg. Ros No. 52 Kel. Maharatu
Kec. Marpoyan Damai Pekanbaru-Riau
Website: al-mujtahadahpress.blogspot.com
E-mail: almujtahadahprss@gmail.com
Hp: 081365662636, 081378712088

ISBN 978-602-9392-82-1



TF-IDF DAN NAÏVE BAYES UNTUK PENCERMINAN DAN MONITORING DISKUSI ONLINE

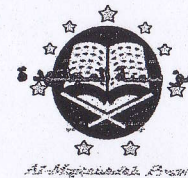


OKFALISA



DR. OKFALISA, MSC

**TF-IDF DAN NAÏVE BAYES
UNTUK FILTERING DAN MONITORING
DISKUSI ONLINE**



TF-IDF DAN NAÏVE BAYES UNTUK FILTERING DAN MONITORING DISKUSI ONLINE

Penulis :

DR. OKFALISA, MSC

Lay Out : Eko Syahputra

Desain Cover: Okfalisa

Percetakan : Utama Grafika

Penerbit :

Al-Mujtahadah Press

Jl. Handayani Gg. Ros No. 52 Kel. Maharatu Kec. Marpoyan Damai

Pekanbaru-Riau Email: almujtahadahprss@gmail.com

Hp. 0813 65662636 Hp. 0813 78712088

Hak Cipta dilindungi Undang-undang. Dilarang memperbanyak sebagian atau keseluruhan isi buku ini dalam bentuk apapun tanpa izin penerbit.

Cetakan :

Pertama, November 2016

© Al-Mujtahadah Press, 2016

Katalog Dalam Terbitan (KDT)

OKFALISA,

TF-IDF DAN NAÏVE BAYES UNTUK FILTERING DAN MONITORING DISKUSI ONLINE/Oleh: Okfalisa,

--Cet. Pekanbaru : Al-Mujtahadah Press, 2016

viii, 65 hlm.; ilus, 21 cm

ISBN 978-602-9392-82-1

1. **Diskusi di Internet**

I. **Judul**

004.693

KATA PENGANTAR

Puji syukur kepada Allah SWT, penulisan buku dengan tema Content Based Spam Filtering pada Diskusi Online dengan Menggunakan Metode TF-IDF dan Naïve Bayes telah selesai dilaksanakan. Buku ini memberikan kontribusi baru terhadap dunia pendidikan dan pembelajaran berbasis ICT melalui proses penjaringan dan ekstraksi informasi dan data pada forum diskusi online. Hal ini memberikan pencerahan pada proses pembelajaran online untuk dapat dimanfaatkan menjadi lebih efektif dan efisien. Melalui aplikasi ini pihak pendidik dapat menilai kemampuan dan karakter siswa melalui tanggapan dan informasi yang diberikan sehingga proses akuisisi pengetahuan dapat terbentuk dengan optimal. Secara analisis penerapan algoritma kedua metode TF-IDF dan Naïve Bayes menggambarkan efektifitas penelusuran yang baik dalam proses filtering resume konten.

Pekanbaru, 25 Agustus 2016

Penulis

DAFTAR ISI

KATA PENGANTAR	iii
DAFTAR ISI	iv
DAFTAR GAMBAR	vi
DAFTAR TABEL	viii

BAB I TEORI DASAR

1.1	Pendahuluan	1
1.1.1	Latar Belakang	1
1.1.2	Batasan Masalah	8
1.1.3	Tujuan	8
1.1.4	Signifikansi	8
1.2	Diskusi Online	10
1.3	Content Based Spam Filtering	11
1.4	Algoritma TF-IDF	12
1.5	Algoritma Naïve Bayes	12
1.6	Penelitian Terkait	13

BAB II METODOLOGI

2.1	Identifikasi Permasalahan	16
2.2	Pembangunan Instrumen	17
2.3	Studi Pendahuluan	18
2.4	Analisa dan Perancangan Model	19
2.5	Implementasi dan Pengujian	20
2.6	Kesimpulan dan Saran	20

BAB III PEMBANGUNAN INSTRUMENT

3.1	Rancangan Instrument dan Penelusuran Proses	21
3.2	Representasi Dokumen/Pembobotan tf-df	30

3.3	Klasifikasi One-Class Naïve Bayes	31
-----	-----------------------------------	----

BAB IV ANALISA DAN PERANCANGAN

4.1	Analisa Sistem	35
4.2	Perancangan Sistem	39
4.3	Context Diagram	39
4.3.1	Data Flow Diagram Level 1	40
4.3.2	Data Flow Diagram Level 2 Proses 2 Preprocessing	41
4.3.3	Data Flow Diagram Level 2 Proses 3 Klasifikasi	42
4.4	ERD (<i>Entity Relation Diagram</i>)	43
4.4.1	Struktur Database	44
4.5	Tampilan Menu Sistem	45

BAB V IMPLEMENTASI DAN PENGUJIAN

5.1	Implementasi	49
5.1.1	Tampilan menu utama system	49
5.2	Pengujian	50
5.2.1	Pengujian akurasi system	50

BAB VI KESIMPULAN DAN SARAN

6.1	Kesimpulan	59
6.2	Saran	60

DAFTAR PUSTAKA	63
----------------	----

DAFTAR GAMBAR

Gambar 2.1. Skema Penelitian	15
Gambar 2.2. Logika Konsep Content Based Spam Filtering	18
Gambar 3.1. Flowchart Klasifikasi Relevansi Komentar dan Topik	22
Gambar 3.2. Preprocessing	23
Gambar 3.3. Contoh Proses <i>Tokenization</i>	24
Gambar 3.4. Contoh Proses <i>Case Folding</i>	25
Gambar 3.5. Flowchart Spelling Normalization	26
Gambar 3.6. Contoh Proses <i>Spelling Normalization</i>	27
Gambar 3.7. Contoh Hasil <i>Filtering</i>	28
Gambar 3.8. Proses <i>Stemming</i> dengan Algoritma Porter Stemmer	29
Gambar 3.9. Proses Klasifikasi dengan <i>One-Class Naive Bayes</i>	32
Gambar 4.1. <i>Context Diagram</i>	39
Gambar 4.2. <i>Data Flow Diagram 1 (DFD) Level 1</i>	40
Gambar 4.3. DFD Level 2 Proses 2 <i>Preprocessing</i>	42
Gambar 4.4. DFD Level 2 Proses 3 Klasifikasi	43
Gambar 4.5. <i>Entity Relation Diagram (ERD)</i>	44
Gambar 4.6. Tampilan Utama Sistem	48
Gambar 5.1. Tampilan Utama Sistem	50
Gambar 5.2. Topik diskusi	51
Gambar 5.3. Komentar Partisipan	51
Gambar 5.4. Ekstraksi dan Klasifikasi Website Indowebster	52
Gambar 5.5. Hasil Ekstraksi Topik Diskusi	52
Gambar 5.6. Hasil Ekstraksi Komentar	53
Gambar 5.7. Hasil Tokenisasi	53

Gambar 5.8. Hasil <i>Stemming</i>	54
Gambar 5.9. Hasil Indeks Topik	54
Gambar 5.10. Hasil Indeks Komentar	55
Gambar 5.11. Hasil Klasifikasi Komentar	55

DAFTAR TABEL

Tabel 3.1 Daftar Kamus Kata Tidak Baku	26
Tabel 3.1 Pembobotan tf-d	30
Tabel 3.2 Nilai Probabilitas Setiap Kata Terhadap Kelas Topik	32
Tabel 3.3 Probabilitas Relevansi Komentar Terhadap Topi	33
Tabel 4.1 Data topik_komenta	44
Tabel 4.2 Data Komentar	45
Tabel 4.3 Data Token	45
Tabel 4.4 Data kamus_kata_tidak_baku	46
Tabel 4.5 Data kamus_stem	46
Tabel 4.6 Data stem	46
Tabel 4.7 Data indeks_top	47
Tabel 4.8 Data indeks kata komentar	48
Tabel 5.1 Matrix Confusion Klasifikasi Website Indowebster	56
Tabel 5.2 Hasil Pengujian	57

BAB I TEORI DASAR

1.1 Pendahuluan

1.1.1. Latar Belakang

Teknologi Information *Communication and Technology* (ICT) sangat erat kaitannya dengan dunia pendidikan, terutama dalam proses belajar mengajar. Pesatnya perkembangan teknologi mendorong para tenaga pendidik (akademisi) untuk lebih memahami dan memanfaatkan teknologi tersebut dalam proses belajar mengajar aktif baik didalam maupun diluar kelas. Hal ini menjadi suatu tantangan yang besar bagi sebuah institusi perguruan tinggi dalam menghadapi kompetisi global dalam memberikan pelayanan pendidikan yang tidak terbatas. Karena kita pahami bersama bahwa diperguruan tinggi inilah transformasi pengetahuan dan komunikasi dapat dibangun tanpa batasan waktu dan tempat, tidak hanya melalui pertemuan *face to face* dikelas saja, namun juga didunia maya.

Internet sebagai salah satu teknologi ICT telah banyak dimanfaatkan untuk melakukan komunikasi dikehidupan sehari-hari. Komunikasi dalam bentuk *face to face* digantikan dengan komunikasi berbasis teknologi. Bahkan banyak sekali media sosial yang dapat dimanfaatkan untuk berkomunikasi, tukar menukar pengalaman, informasi dan pengetahuan (Murray, 2008). Beberapa sosial media yang cukup digemari dalam melakukan komunikasi virtual antara lain adalah Facebook, Linked, Twitter, Youtube, MySpace, Google Plus, Classroom 2.0, Plurk, etc. Bahkan pengguna sosial media ini setiap tahunnya terus meningkat, terutama