



**PENERAPAN ALGORITMA *TERM FREQUENCY-
INVERSE DOCUMENT FREQUENCY* (TF-IDF) DALAM
RANCANG BANGUN *WEBSITE* PENCARIAN STANDAR
NASIONAL INDONESIA**

Nickolas Junior 41819120008

Muhamad Doni Munggaran 41819120057

Desi Yosephine Sinaga 41819110059

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022**



PENERAPAN ALGORITMA *TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY* (TF-IDF) DALAM RANCANG BANGUN *WEBSITE* PENCARIAN STANDAR NASIONAL INDONESIA

Laporan Tugas Akhir

Diajukan Untuk Melengkapi Salah
Satu Syarat
Memperoleh Gelar Sarjana
Komputer

Oleh:

Nickolas Junior 41819120008

Muhamad Doni Munggaran 41819120057

Desi Yosephine Sinaga 41819110059

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

Nama Mahasiswa (1) : Nickolas Junior

NIM 41819120008

Nama Mahasiswa (2) : Muhamad Doni Munggaran

NIM 41819120057

Nama Mahasiswa (3) : Desi Yosephine Sinaga

NIM 41819110059

Judul Tugas Akhir : Penerapan Algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) Dalam Rancang Bangun *Website* Pencarian Standar Nasional Indonesia

Menyatakan bahwa Laporan Tugas Akhir ini adalah hasil karya nama yang tercantum diatas dan bukan plagiat. Apabila ternyata ditemukan di dalam laporan Tugas Akhir ini terdapat unsur plagiat, maka nama diatas siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 7 Februari 2022



Nickolas Junior

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa (1) : Nickolas Junior
 NIM : 41819120008
 Nama Mahasiswa (2) : Muhamad Doni Munggaran
 NIM : 41819120057
 Nama Mahasiswa (3) : Desi Yosephine Sinaga
 NIM : 41819110059
 Judul Tugas Akhir : Penerapan Algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) Dalam Rancang Bangun *Website* Pencarian Standar Nasional Indonesia

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 7 Februari 2022



Nickolas Junior


LEMBAR PENGESAHAN

Nama Mahasiswa (1) : Nickolas Junior 41819120008
NIM
Nama Mahasiswa (2) : Muhamad Doni Munggaran
NIM 41819120057
Nama Mahasiswa (3) : Desi Yosephine Sinaga
NIM 41819110059
Judul Tugas Akhir : Penerapan Algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) Dalam Rancang Bangun *Website* Pencarian Standar Nasional Indonesia

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.


Jakarta, 7 Februari 2022

Menyetujui,


(Nur Ani, ST. MMSI)

Dosen Pembimbing

Mengetahui,


(Yunita Sartika Sari, S.Kom., M.Kom)
Prodi Sistem Informasi


(Ratna Mutu Manikam, S.Kom., M.T) Sek.
Ka.Prodi Sistem Informasi

LEMBAR PERSETUJUAN

Nama Mahasiswa (1) : Nickolas Junior
NIM : 41819120008
Nama Mahasiswa (2) : Muhamad Doni Munggaran
NIM : 41819120057
Nama Mahasiswa (3) : Desi Yosephine Sinaga
NIM : 41819110059
Judul Tugas Akhir : PENERAPAN ALGORITMA TERM FREQUENCY-
INVERSE DOCUMENT FREQUENCY (TD-IDF) DALAM RANCANG
BANGUN WEBSITE PENCARIAN STANDAR NASIONAL INDONESIA

Tugas Akhir ini telah diperiksa dan disetujui,
Jakarta, 7 Februari 2022

Menyetujui,

UNIVERSITAS
MERCUBUANA

(Nur Ani, ST., MMSI)

Dosen pembimbing

KATA PENGANTAR

Puji syukur kita panjatkan kepada Tuhan Yang Maha Esa, karena berkat rahmat dan karunia-Nya kami akhirnya mampu menyelesaikan Laporan Tugas Akhir “Penerapan Algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) Dalam Rancang Bangun *Website* Pencarian Standar Nasional Indonesia” dengan baik dan tepat waktu.

Penulis menyadari bahwa tanpa bantuan, bimbingan serta doa, laporan ini mungkin tidak dapat diselesaikan dengan baik. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Tuhan Yang Maha Esa, yang selalu memberikan kekuatan dan semangat baru setiap akhirnya untuk terus berproses setiap harinya.
2. Ibu Nur Ani, ST. MMSI, selaku dosen pembimbing tugas akhir, yang selama ini memberikan arahan dan serta meluangkan waktu serta bimbingan dalam penyusunan laporan.
3. Teman satu tim yang bekerjasama dengan baik selama proses pembuatan *website* dan laporan serta senantiasa tetap aktif untuk melakukan perbaikan.
4. Orang Tua dan Teman yang selalu memberikan motivasi dan doa selama penyusunan laporan.
5. Penulis mengucapkan terima kasih kepada semua pihak yang pada penulisan penelitian ini kami sadur dalam penulisan laporan.

Kami mendapati banyak kesulitan dalam penyelesaiannya, namun segala usaha serta doa sehingga penyusunan laporan ini dapat terselesaikan dengan baik. Kami pun menyadari dalam penyusunan laporan ini masih banyak kekurangan dan jauh dari kesempurnaan hingga kami mengharapkan saran dan kritik yang membangun agar dapat berguna dan bermanfaat bagi para pembaca dan pihak-pihak lain yang berkepentingan.

Jakarta, 7 Februari 2022

Penulis

DAFTAR ISI

SURAT PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR.....	iii
LEMBAR PENGESAHAN	iv
LEMBAR PERSETUJUAN	v
ABSTRAK	vi
<i>ABSTRACT</i>	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan dan Manfaat Penelitian.....	3
1.4.1 Tujuan Penelitian.....	3
1.4.2 Manfaat Penelitian.....	3
1.5 Sistematika Penulisan	3
BAB II.....	5
TINJAUAN PUSTAKA.....	5
2.1 Kementerian Perindustrian	5
2.1.1 Standard Nasional Indonesia (SNI).....	6
2.2 <i>Python</i>	7
2.3 <i>Web Crawler</i>	7
2.4 <i>String Matching</i>	8
2.5 <i>TF-IDF (Term Frequency — Inverse Document Frequency)</i>	8
2.6 Text Preprocessing.....	10
2.7 Penelitian Terdahulu	10
BAB III.....	15
METODE PENELITIAN.....	15
3.1 Lokasi Penelitian	15
3.2 Sarana Pendukung.....	15
3.3 Teknik Pengumpulan Data	16
3.4 Metode Pengembangan <i>Website</i>	17
3.4.1 Penjelasan Tahap-tahap dalam pengembangan Model <i>Prototype</i>	17
BAB IV	19
HASIL DAN PEMBAHASAN	19

4.1	Analisis Masalah dan Analisis Kebutuhan	19
4.1.1	Analisis Masalah	19
4.1.2	Analisis Kebutuhan.....	20
4.2	Analisis Proses Bisnis.....	22
4.3	Algoritma TF-IDF dalam proses pencarian kata kunci	23
4.3.1	<i>Text preprocessing</i>	23
4.3.2	Algoritma TF-IDF.....	24
4.4	<i>Crawling Data</i>	28
4.5	Rancangan <i>Website</i> Pencarian SNI	32
4.5.1	<i>Usecase Diagram</i>	32
4.5.2	<i>Activity Diagram</i>	33
4.5.3	<i>Class Diagram & Rancangan Database</i>	34
4.5.4	Rancangan Tampilan.....	36
4.6	Perbandingan <i>website</i> lama dan <i>website</i> pencarian SNI.....	38
BAB V	39
KESIMPULAN DAN SARAN	39
5.1	Kesimpulan.....	39
5.2	Saran	39
DAFTAR PUSTAKA	40
LAMPIRAN	41



DAFTAR TABEL

Tabel 1 Penelitian terkait 1.....	11
Tabel 2 Penelitian terkait 2.....	11
Tabel 3 Penelitian terkait 3.....	12
Tabel 4 Penelitian terkait 4.....	13
Tabel 5 Penelitian terkait 5.....	13
Tabel 6 Tabel Abstraksi	25
Tabel 7 Tabel proses <i>Tokenizing</i>	26
Tabel 8 Tabel setelah <i>steaming</i> dan <i>filtering</i>	26
Tabel 9 Tabel Perhitungan	27
Tabel 10 <i>mst_berita</i>	35
Tabel 11 <i>stop_word</i>	35
Tabel 12 <i>document</i>	35
Tabel 13 <i>contact_us</i>	35
Tabel 14 <i>freq</i>	36
Tabel 15 <i>admin</i>	36
Tabel 16 Tabel Perbandingan <i>Website</i>	38



DAFTAR GAMBAR

Gambar 1 Model <i>Prototype</i>	17
Gambar 2 <i>Diagram</i> analisis <i>fishbone</i>	19
Gambar 3 Tampilan <i>Website</i> SNI Pustand Industri.....	21
Gambar 4 Tampilan <i>website</i> SNI milik BSN	21
Gambar 5 <i>Flow</i> proses <i>tokenizing</i>	23
Gambar 6 <i>Flow</i> proses <i>Filtering</i>	24
Gambar 7 <i>Flowchart</i> proses <i>crawling</i> berita.....	29
Gambar 8 <i>source code crawling</i> berita.....	30
Gambar 9 <i>source code input</i> data berita.....	30
Gambar 10 <i>Flowchart</i> proses <i>crawling</i> dokumen SNI.....	31
Gambar 11 <i>Usecase diagram Website</i> Pencarian SNI	32
Gambar 12 <i>Activity Diagram</i>	33
Gambar 13 <i>Class Diagram</i>	34
Gambar 14 Rancangan Tampilan Menu Data Berita	36
Gambar 15 Rancangan <i>Contact Us</i>	37
Gambar 16 Rancangan Tampilan Menu Pencarian SNI	37
Gambar 17 Rancangan Tampilan Hasil Pencarian.....	37

