



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

Muhamad Doni Munggaran 41819120057

Nickolas Junior 41819120008

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:

Muhamad Doni Munggaran 41819120057

Nickolas Junior 41819120008

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) : Muhamad Doni Munggaran  
NIM : 41819120057  
Nama Mahasiswa (2) : Desi Yosephine Sinaga  
NIM : 41819110059  
Nama Mahasiswa (3) : Nickolas Junior  
NIM : 41819120008  
Judul Tugas Akhir : PENERAPAN ALGORITMA TERM FREQUENCY-  
INVERSE DOCUMENT FREQUENCY (TD-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.

UNIVERSITAS  
MERCU BUANA

Jakarta, 7 Februari 2022



Desi Yosephine Sinaga

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa (1) : Muhamad Doni Munggaran  
NIM : 41819120057  
Nama Mahasiswa (2) : Desi Yosephine Sinaga  
NIM : 41819110059  
Nama Mahasiswa (3) : Nickolas Junior  
NIM : 41819120008  
Judul Tugas Akhir : PENERAPAN ALGORITMA TERM FREQUENCY-  
INVERSE DOCUMENT FREQUENCY (TD-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



Desi Yosephine Sinaga

## LEMBAR PERSETUJUAN

Nama Mahasiswa (1) : Muhamad Doni Munggaran  
NIM : 41819120057  
Nama Mahasiswa (2) : Desi Yosephine Sinaga  
NIM : 41819110059  
Nama Mahasiswa (3) : Nickolas Junior  
NIM : 41819120008  
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,



(Nur Ani, ST., MMSI)

UNIVERSITAS  
MERCU BUANA

## LEMBAR PENGESAHAN

Nama Mahasiswa (1) : Muhamad Doni Munggaran  
NIM : 41819120057  
Nama Mahasiswa (2) : Desi Yosephine Sinaga  
NIM : 41819110059  
Nama Mahasiswa (3) : Nickolas Junior  
NIM : 41819120008  
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, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 7 Februari 2022

Menyetujui,

(Nur Ani, ST, MMSI)

Dosen Pembimbing

UNIVERSITAS  
MERCU BUANA

Mengetahui,



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



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

## 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 mambangun agar dapat berguna dan bermanfaat bagi para pembaca dan pihak-pihak lain yang berkepentingan.

Jakarta, 7 Februari 2022

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
SURAT PERNYATAAN ORISINALITAS .....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR.....	iii
LEMBAR PERSETUJUAN .....	iv
LEMBAR PENGESAHAN .....	v
ABSTRAK .....	vi
<i>ABSTRACT</i> .....	vii
KATA PENGANTAR.....	viii
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 TF-IDF ( <i>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
<b>BAB IV</b>	.....	<b>19</b>
<b>HASIL DAN PEMBAHASAN</b>	.....	<b>19</b>
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
4.5.4	Rancangan Tampilan .....	36
4.6	Perbandingan <i>website</i> lama dan <i>website</i> pencarian SNI.....	38
<b>BAB V</b>	.....	<b>39</b>
<b>KESIMPULAN DAN SARAN</b>	.....	<b>39</b>
5.1	Kesimpulan .....	39
5.2	Saran.....	39
<b>DAFTAR PUSTAKA</b>	.....	<b>40</b>
<b>LAMPIRAN</b>	.....	<b>41</b>



## DAFTAR TABEL

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



## DAFTAR GAMBAR

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

