

**IMPLEMENTASI *TEXT MINING* UNTUK ANALISIS
SENTIMEN PROTOKOL KESEHATAN DI INDONESIA
MENGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR***



Skripsi

Untuk memenuhi syarat memperoleh Derajat
Sarjana Teknik (S.T.)

Oleh:

FIDYA FARASALSABILA

NIM 180155201014

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MARITIM RAJA ALI HAJI
TANJUNGPINANG
2022**

**IMPLEMENTASI *TEXT MINING* UNTUK ANALISIS
SENTIMEN PROTOKOL KESEHATAN DI INDONESIA
MENGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR***



Skripsi

Untuk memenuhi syarat memperoleh derajat
Sarjana Teknik (S.T.)

Oleh:

FIDYA FARASALSABILA

NIM 180155201014

Telah mengetahui dan disetujui oleh :

Pembimbing I,

Nola Ritha, S.T., M.Cs.
NIP. 199011142019032016

Pembimbing II,

Nurul Hayaty, S.T., M.Cs.
NIP. 199103272019032019

HALAMAN PENGESAHAN

Judul : Implementasi *Text Mining* untuk Analisis Sentimen Protokol Kesehatan di Indonesia Menggunakan Algoritma *K-Nearest Neighbor*

Nama : Fidyfa Farasalsabila

NIM : 180155201014

Jurusan : Teknik Informatika

telah dipertahankan di depan Dewan Penguji dan dinyatakan lulus pada tanggal 14 Juli 2022

Susunan Tim Pembimbing dan Penguji

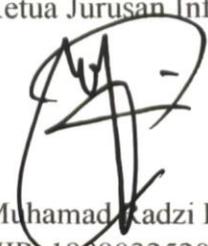
Jabatan	Nama	Tanda Tangan	Tanggal
Pembimbing I	: Nola Ritha, S.T., M.Cs.		26 Juli 2022
Pembimbing II	: Nurul Hayaty, S.T., M.Cs.		26 Juli 2022
Ketua Penguji	: Nola Ritha, S.T., M.Cs.		26 Juli 2022
Anggota	: 1. Martaleli Bettiza, S.Si., M.Sc		26 Juli 2022
	2. Tekad Matulatan, S.Sos., S.Kom., M.Inf.Tech		26 Juli 2022
	3. Muhamad Radzi Rathomi, S.Kom., M.Cs		26 Juli 2022

Tanjungpinang, 26 Juli 2022

Universitas Maritim Raja Ali Haji

Fakultas Teknik

Ketua Jurusan Informatika,


Muhamad Radzi Rathomi, S.Kom., M.Cs

NIP. 198903252019031014

PERNYATAAN ORISINALITAS

Dengan ini saya menyatakan bahwa skripsi saya yang berjudul Implementasi *Text Mining* untuk Analisis Sentimen Protokol Kesehatan di Indonesia menggunakan Algoritma *K-Nearest Neighbor* adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Jika kemudian hari ternyata terbukti pernyataan saya ini tidak benar dan melanggar peraturan yang sah dalam karya tulis dan hak intelektual maka saya bersedia ijazah yang telah saya terima untuk ditarik kembali oleh Universitas Maritim Raja Ali Haji.

Tanjungpinang, 14 Juli 2022

Yang menyatakan



Fidya Farasalsabila
NIM. 180155201014

HALAMAN PERSEMBAHAN

Karya tulis ini saya persembahkan untuk

Bapak dan Mamak saya.

Terima kasih sudah menyayangi, mendukung, memfasilitasi dan mengayomi.

Kebahagiaan untukmu.



HALAMAN MOTO

“Your life is as good as your mindset”

-fidid



KATA PENGANTAR

Alhamdulillahirobbil'alamin puji dan syukur penulis ucapkan kehadirat Allah SWT karena berkat kasih dan karuniaNya penulis dapat menyelesaikan skripsi yang berjudul “Implementasi *Text Mining* untuk Analisis Sentimen Protokol Kesehatan di Indonesia menggunakan Algoritma *K-Nearest Neighbor*”. Penulis menulis skripsi ini guna memenuhi salah satu persyaratan yang ditetapkan untuk memperoleh gelar sarjana Strata-1 di Jurusan Teknik Informatika Fakultas Teknik Universitas Maritim Raja Ali Haji.

Penulis menyadari bahwa tanpa adanya bimbingan dan dorongan dari berbagai pihak maka penulisan skripsi ini tidak berjalan dengan lancar. Oleh karena itu pada kesempatan ini, penulis ingin menyampaikan terima kasih kepada:

1. Bapak Sapta Nugraha, S.T., M.Eng, selaku Dekan Fakultas Teknik Universitas Maritim Raja Ali Haji
2. Bapak Muhamad Radzi Rathomi, S.Kom., M.Cs, selaku Kepala Jurusan Program Studi Teknik Informatika.
3. Bapak Tekad Matulatan, S.Sos., S.Kom., M.Inf.Tech, selaku dosen Pembimbing Akademik (PA) yang telah memberikan semangat dan arahan.
4. Ibu Nola Ritha, S.T., M.Cs, selaku Dosen Pembimbing I yang telah menyediakan waktu, tenaga dan pikiran serta membimbing dalam penyusunan skripsi ini.
5. Ibu Nurul Hayaty, S.T., M.Cs, selaku Pembimbing II yang telah membimbing penulis dan memberikan masukan serta mengarahkan dalam penyelesaian skripsi.
6. Para Dosen Fakultas Teknik yang telah banyak membantuk membekali ilmu dan berbagi pengalaman selama masa perkuliahan, serta staff yang selalu dengan senang hati melayani dalam administrasi perkuliahan.
7. Kedua orang tua penulis tercinta, Ibu Nuraini dan Bapak Supardi, S.Pd., MM. yang selalu menyebutkan nama penulis dalam setiap do'a. Terimakasih Ibu, Bapak.
8. Teman-teman Teknik Informatika angkatan 2018 (TI'18), yang tidak mungkin saya tulis semua satu-satu karena nantinya akan menjadi seperti *datasets crawling* penelitian saya. Terimakasih telah menuliskan kisah indah masa kuliah 4 tahun ini.

Penulis menerima kritik dan saran yang membangun agar menjadi perbaikan bagi penulis dikemudian hari. Semoga Allah mengampuni setiap dosa yang kita lakukan, dan mohon maaf atas kesalahan yang dilakukan.

Billahittaufiq walhidayah,

Wassalamualaikum warrahmatullah wabarakatuh.

Tanjungpinang, 14 Juli 2022

Penulis,



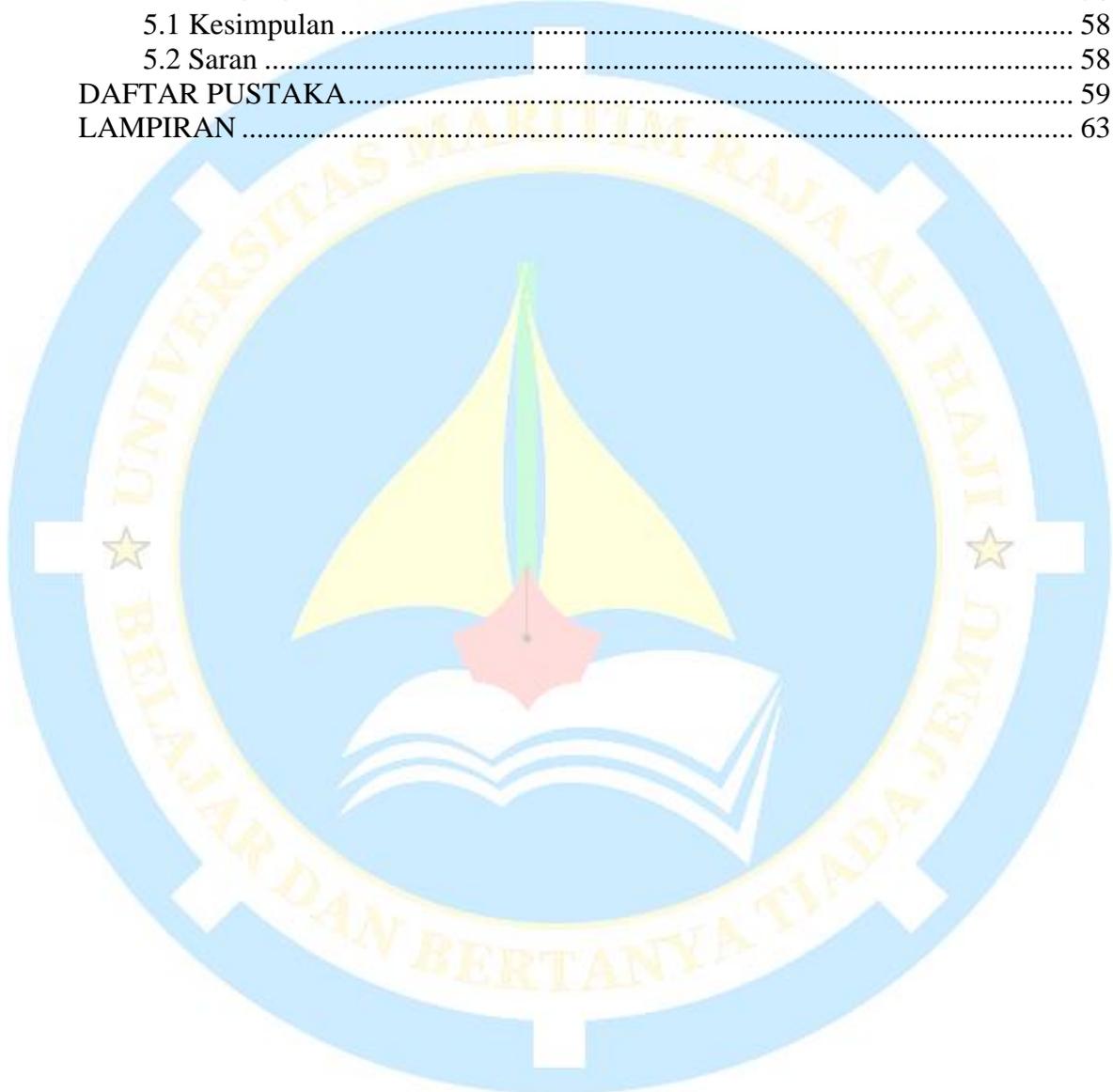
Fidya Farasalsabila



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
PERNYATAAN ORISINALITAS	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
ABSTRAK	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II KAJIAN LITERATUR	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	7
2.2.1 Analisis Sentimen	7
2.2.2 Protokol Kesehatan	8
2.2.3 <i>Twitter</i>	9
2.2.4 <i>Phyton</i>	9
2.2.5 <i>Jupyter Notebook</i>	11
2.2.6 <i>Crawling</i>	12
2.2.7 VADER (Valence Aware Dictionary for Sentiment Reasoning)	12
2.2.8 <i>Text Mining</i>	13
2.2.9 TF-IDF (<i>Term Frequency–Inverse Document Frequency</i>)	15
2.2.10 <i>K-Nearest Neighbor</i>	16
2.2.11 <i>Cosine Similarity</i>	17
2.2.12 <i>Confusion Matrix</i>	18
2.2.13 <i>Database</i>	19
BAB III METODE PENELITIAN	21
3.1 Waktu Penelitian	21
3.2 Jenis Penelitian	21
3.3 Materi dan Bahan Penelitian	21
3.4 Alat dan Instrumen Penelitian	24
3.5 Jadwal Penelitian	24
3.6 Prosedur Penelitian	25
3.7 Perhitungan Manual	25
3.7.1 Pre-processing	26
3.7.2 Proses Pembobotan Kata TF-IDF	37
3.7.3 Implementasi K-Nearest Neighbors	40

3.8 Analisis Perancangan Sistem	44
3.8.1 Data Flow Diagram Level 0.....	44
3.8.2 Data Flow Diagram Level 1.....	45
3.8.3 Entity Relationship Diagram (ERD).....	47
3.8.4 Tampilan Antarmuka Sistem	47
BAB IV HASIL DAN PEMBAHASAN.....	55
4.1 Hasil Pengujian	55
4.2 Pembahasan.....	57
BAB V PENUTUP	58
5.1 Kesimpulan	58
5.2 Saran	58
DAFTAR PUSTAKA.....	59
LAMPIRAN	63



DAFTAR TABEL

Tabel 2.1 Tabel Confusion Matrix	18
Tabel 2.2 Tabel Confusion Matrix Multiclass	18
Tabel 3.1 Sampel Data Latih.....	22
Tabel 3.2 Sampel Data Uji	23
Tabel 3.3 Spesifikasi Laptop.....	24
Tabel 3.4 Jadwal Penelitian.....	24
Tabel 3.5 Tokenizing.....	26
Tabel 3.6 Stop word removal	29
Tabel 3.7 Stemming dan Lemmatization	33
Tabel 3.8 Hasil Pre-processing Data	35
Tabel 3.9 Perhitungan TF.....	37
Tabel 3.10 Perhitungan TF Normalisasi	38
Tabel 3.11 Perhitungan IDF.....	39
Tabel 3.12 Perhitungan TF-IDF	39
Tabel 3.13 Hasil Perkalian Skalar antara Data Uji dengan Data Latih	41
Tabel 3.14 Hasil Perhitungan Panjang Dokumen	42
Tabel 3.15 Hasil Perhitungan Jarak.....	43
Tabel 3.16 Hasil Perangkingan Dokumen.....	43
Tabel 3.17 Dokumen Ranking Teratas.....	44
Tabel 4.1 Hasil Uji Akurasi Nilai K.....	55
Tabel 4.2 Tabel Confusion Matrix Data Uji K=4	55

DAFTAR GAMBAR

Gambar 2.1 Logo <i>Twitter</i>	9
Gambar 2.2 Logo <i>Python</i> (<i>Sumber:https://www.python.org/</i>).....	10
Gambar 2.3 Logo <i>Jupyter Notebook</i> (<i>Sumber:https://jupyter.org/</i>)	11
Gambar 2.4 Contoh Klasifikasi KNN (<i>Sumber: Budiharto, 2016</i>)	17
Gambar 3.1 Data hasil <i>crawling</i>	22
Gambar 3.2 Data Flow Diagram Level 0.....	45
Gambar 3.3 Data Flow Diagram Level 1.....	46
Gambar 3.4 Entity Relationship Diagram (ERD).....	47
Gambar 3.5 Tampilan Halaman Login	48
Gambar 3.6 Tampilan Halaman Beranda	48
Gambar 3.7 Tampilan Halaman Data Latih.....	49
Gambar 3.8 Tampilan Halaman Data Uji	50
Gambar 3.9 Tampilan Halaman Hasil Preprocessing.....	51
Gambar 3.10 Tampilan Halaman Proses Klasifikasi K-Nearest Neighbor	52
Gambar 3.11 Tampilan Halaman Hasil Klasifikasi K-Nearest Neighbor	53
Gambar 3.12 Tampilan Halaman Hasil TF-IDF	54
Gambar 3.13 Tampilan Halaman Logout	54
Gambar 4.1 Visualisasi Uji Akurasi K=4	70
Gambar 4.2 Visualisasi Uji Akurasi K Keseluruhan	70