

TUGAS AKHIR

**KOMBINASI ALGORITMA SUPPORT VECTOR
MACHINE DAN LEXICON BASED PADA ANALISIS
SENTIMEN TWITTER**

(Studi Data: PP Tapera No 25 Tahun 2020)



RINDU HAFIL MUHAMMADI

NIM. 17102115

**PROGRAM STUDI S1 INFORMATIKA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO
2021**

TUGAS AKHIR

**KOMBINASI ALGORITMA SUPPORT VECTOR
MACHINE DAN LEXICON BASED PADA ANALISIS
SENTIMEN TWITTER**

(Studi Data: PP Tapera No 25 Tahun 2020)

**COMBINATION OF SUPPORT VECTOR MACHINE
AND LEXICON BASED ALGORITHM IN TWITTER
SENTIMENT ANALYSIS**

(Data Study: PP Tapera No. 25 Tahun 2020)

Disusun Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer



RINDU HAFIL MUHAMMADI

NIM. 17102115

PROGRAM STUDI S1 INFORMATIKA

FAKULTAS INFORMATIKA

INSTITUT TEKNOLOGI TELKOM PURWOKERTO

2021

LEMBAR PENGESAHAN PEMBIMBING

**KOMBINASI ALGORITMA SUPPORT VECTOR
MACHINE DAN LEXICON BASED PADA ANALISIS
SENTIMEN TWITTER**

(Studi Data: PP Tapera No 25 Tahun 2020)

**COMBINATION OF SUPPORT VECTOR MACHINE
AND LEXICON BASED ALGORITHM IN TWITTER
SENTIMENT ANALYSIS**

(Data Study: PP Tapera No. 25 Tahun 2020)

Dipersiapkan dan Disusun Oleh:

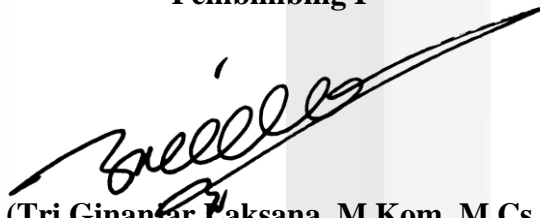
RINDU HAFIL MUHAMMADI

NIM. 17102115

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tugas Akhir
Pada hari Jum'at, 27 Agustus 2021

Pembimbing I

Pembimbing II



(Tri Ginanjar Laksana, M.Kom.,M.Cs.) (Amalia Beladinna Arifa, S.Pd., M.Cs)

NIDN. 0407088502



NIDN. 0606019201

Tugas Akhir ini diterima sebagai salah satu persyaratan untuk
memperoleh gelar Sarjana Komputer

Tanggal ...

Kaprodi,

-

(Auliya Burhanuddin, S.Si., M. Kom.,)

NIDN. 0630058202

LEMBAR PENETAPAN PENGUJI
KOMBINASI ALGORITMA SUPPORT VECTOR
MACHINE DAN LEXICON BASED PADA ANALISIS
SENTIMEN TWITTER
(Studi Kasus: PP Tapera No 25 Tahun 2020)

COMBINATION OF SUPPORT VECTOR MACHINE
AND LEXICON BASED ALGORITHM IN TWITTER
SENTIMENT ANALYSIS
(Data Study: PP Tapera No. 25 2020)

Dipersiapkan dan Disusun Oleh:

RINDU HAFIL MUHAMMADI

NIM. 17102115


Tugas Akhir Telah diuji dan Dinilai Panitia Penguji
Program Studi Informatika
Fakultas Informatika
Institut Teknologi Telkom Purwokerto
Pada Tanggal: 27 Agustus 2021


Ketua Penguji


(Tri Ginanjar Laksana, M.Kom.,M.Cs.)
NIDN. 0407088502

Penguji 1

Penguji II


(Agi Prasetiadi, S.T., M.Eng.)
NIDN. 0617098802


(Siti Khomsah, S.Kom.,M.Cs.)
NIDN. 0517108101

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertandatangan di bawah ini,

Nama Mahasiswa : Rindu Hafil Muhammadi
NIM : 17102115
Program Studi : Teknik Informatika

Menyatakan bahwa Tugas Akhir dengan judul berikut:

KOMBINASI ALGORITMA SUPPORT VECTOR MACHINE DAN LEXICON BASED PADA ANALISIS SENTIMEN TWITTER (Studi Data: PP Tapera No 25 Tahun 2020)

Dosen Pembimbing Utama : Tri Ginanjar Laksana, M.Kom.,M.Cs.

Dosen Pembimbing Pendamping : Amalia Beladonna Arifa, S.Pd., M.Cs

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Institut Teknologi Telkom Purwokerto maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan, dan penelitian Saya Sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing.
3. Dalam Karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab saya, bukan tanggungjawab Institut Teknologi Telkom Purwokerto.
5. Pernyataan ini Saya buat dengan sesungguhnya, apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima Sanksi Akademik dengan pencabutan gelar yang sudah diperoleh serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Tegal, 15 Agustus 2021

Yang Menyatakan,



Rindu Hafil Muhammadi

KATA PENGANTAR

Segala puji bagi Allah SWT yang telah memberikan rahmat dan karunia Nya kepada penulis, sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik. Shalawat dan salam senantiasa tercurah kepada Rasulullah SAW. Penyusunan tugas akhir ini dimaksudkan untuk memenuhi sebagian syarat-syarat guna mencapai gelar Sarjana Komputer, Institut Teknologi Telkom Purwokerto.

Penulis menyadari bahwa penulisan ini tidak dapat terselesaikan tanpa dukungan dari berbagai pihak baik moril maupun materil. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini terutama kepada:

1. Bapak Sugiarto S.Pd dan Ibu Nurhayati, selaku kedua orang tua penulis, yang selalu memberikan dukungan baik moril maupun materil serta doa yang tiada henti-hentinya kepada penulis.
2. Saudari Maulida Rizaq Istiqomah selaku kakak dari penulis, yang telah memberikan dukungan, dan supportnya kepada penulis.
3. Saudari Shabrina Nurul Faza, selaku adik dari penulis, yang telah memberikan dukungan, dan supportnya kepada penulis.
4. Bapak/Ibu Dekan Fakultas Informatika, Institut Teknologi Telkom Purwokerto.
5. Bapak/Ibu Ketua Jurusan Informatika, Institut Teknologi Telkom Purwokerto.
6. Bapak Tri Ginanjar Laksana, S.Kom., M.Cs. selaku dosen Pembimbing I yang telah bersedia membimbing dan mengarahkan penulis selama menyusun tugas akhir dan memberikan banyak ilmu serta solusi pada setiap permasalahan atas kesulitan dalam penulisan tugas akhir ini.
7. Ibu Amalia Beladonna Arifa, S.Pd., M.Cs. selaku dosen Pembimbing II yang telah bersedia membimbing dan mengarahkan penulis selama menyusun tugas akhir dan memberikan banyak ilmu serta solusi pada setiap permasalahan atas kesulitan dalam penulisan tugas akhir ini.

8. Bapak/Ibu Dosen Fakultas Informatika yang telah memberikan pengetahuan yang sangat bermanfaat selama masa perkuliahan.
9. Seluruh rekan-rekan saya di HMIF Periode 2018/2019, HMIF Periode 2019/2020, UKM Manggala Periode 2019/2020 yang telah membantu saya di Organisasi Mahasiswa.
10. Seluruh teman teman dekat saya yang selalu mensupport saya yaitu Mochammad Iqbal Fatria, Faqih Abdillah, Nathanael Novarian, Fadil Al Afgani, Ilham Fadhillah Akbar, Rayhan Hidayat, dan Dimas Purwita Sari.
11. Dan Saudari Intan Rizki Arinta Putri yang selalu memberikan doa, semangat dan dukungannya kepada penulis.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna dikarenakan terbatasnya pengalaman dan pengetahuan yang dimiliki penulis. Oleh karena itu, penulis mengharapkan segala bentuk saran serta masukan bahkan kritik yang membangun. Semoga tugas akhir ini dapat bermanfaat bagi para pembaca dan semua pihak khususnya dalam bidang Informatika.

Tegal, 15 Agustus 2021

Penulis,

(Rindu Hafil Muhammadi)

DAFTAR ISI

LEMBAR PENGESAHAN PEMBIMBING	iii
LEMBAR PENETAPAN PENGUJI	iv
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
DAFTAR SINGKATAN	xiv
DAFTAR LAMPIRAN	xv
INTISARI	xvi
ABSTRACT	xvii
BAB I	1
PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Perumusan Masalah	4
1.3. Pertanyaan Penelitian	4
1.4. Tujuan Penelitian	5
1.5. Batasan Masalah	5
1.6. Manfaat Penelitian	6
BAB II	7
TINJAUAN PUSTAKA	7
2.1. Penelitian Sebelumnya	7
2.2. Dasar Teori	20
2.2.1. PP 25 Tahun 2020	20
2.2.2. Twitter	21
2.2.3. Klasifikasi	22
2.2.4. Analisis Sentimen	22
2.2.5. Text Preprocessing	23
2.2.6. Lexicon Based	23

2.2.7. Pembobotan TF-IDF	24
2.2.8. Support Vector Machine	25
2.2.9. Metode Grid Search	26
2.2.10. Confusion Matrix	27
2.2.11. Scikit-Learn	28
BAB III	30
METODOLOGI PENELITIAN	30
3.1. Subjek dan Objek Penelitian	30
3.2. Alat dan Bahan	30
3.2.1. Alat	30
3.2.2. Bahan	30
3.3. Tahapan Penelitian	31
3.3.1. Studi Pustaka	32
3.3.2. Perumusan Masalah dan Tujuan Penelitian	32
3.3.3. Teknik Pengumpulan Data	32
3.3.4. Analisa Data	34
3.4. Pengolahan Data	36
3.4.1. Preprocessing	39
3.4.2. Pelabelan Data	40
3.4.3. Pembobotan Term	43
3.4.4. Pengujian Support Vector Machine	45
3.4.5. Evaluasi Confusion Matrix	46
3.5. Analisa Hasil Pengolahan Data	47
3.6. Kesimpulan	47
BAB IV	48
HASIL DAN PEMBAHASAN	48
4.1. Hasil Pengolahan Data	48
4.1.1. Hasil Preprocessing	48
4.1.2. Hasil Pelabelan	50
4.1.3. Hasil Pembobotan TFIDF	52
4.1.4. Hasil Pengujian Support Vector Machine	55

4.2. Analisis	63
4.2.1. Perbandingan Pengujian Kondisi Unbalanced Dataset dan Kondisi Balanced Dataset	65
4.2.2. Perbandingan Pengujian Penggunaan Label Manual dan Lexicon Based dalam Proses Kombinasi	67
BAB V	71
KESIMPULAN DAN SARAN	71
5.1. Kesimpulan	71
5.2. Saran	72
DAFTAR PUSTAKA	73
LAMPIRAN	79
Lampiran 1. Perhitungan Hasil Confusion Matrix Tiap Kernel	79
Lampiran 2. Source Code Pengujian	81
Lampiran 3. Source Code Perbandingan Pengujian	90

DAFTAR TABEL

Tabel 2. 1 Perbandingan Referensi	10
Tabel 3. 1 Contoh Dataset Penelitian	33
Tabel 3. 2 Analisis Data	34
Tabel 3. 3 Kata Slang pada Data	35
Tabel 3. 4 Contoh Kamus Inset Positif dan Negatif	40
Tabel 3. 5 Contoh Kamus Swear Word	41
Tabel 3. 6 Contoh Kamus Sentiment Word	41
Tabel 3. 7 Contoh Hasil Pelabelan Lexicon Based	43
Tabel 3. 8 Persamaan Kernel Pada SVM	45
Tabel 4. 1 Hasil Casefolding	48
Tabel 4. 2 Hasil Cleansing	48
Tabel 4. 3 Hasil Tokenizing	49
Tabel 4. 4 Hasil Slangwords	49
Tabel 4. 5 Hasil Stopwords	49
Tabel 4. 6 Hasil Stemming	49
Tabel 4. 7 Contoh Dataset Hasil Preprocessing	50
Tabel 4. 8 Contoh Hasil Pelabelan Dataset dengan Lexicon	51
Tabel 4. 9 Distribusi Label Tweet	52
Tabel 4. 10 Distribusi Pembagian Dataset	53
Tabel 4. 11 Hasil TF-IDF Data Train	53
Tabel 4. 12 Hasil TF-IDF Data Test	54
Tabel 4. 13 Range Hyperparameter Kernel Linear	56
Tabel 4. 14 Nilai Parameter Terbaik Tiap Kernel	57
Tabel 4. 15 Confusion Matrix Kernel Linear	58
Tabel 4. 16 Confusion Matrix Kernel RBF	58
Tabel 4. 17 Confusion Matrix Kernel Polynomial	59
Tabel 4. 18 Confusion Matrix Kernel Sigmoid	59
Tabel 4. 19 Hasil Pengujian Tiap Kernel	60
Tabel 4. 20 Contoh Hasil Klasifikasi	61

Tabel 4. 21 Perbedaan Hasil Prediktif dan Aktual	61
Tabel 4. 22 Distribusi Label Dataset untuk Pengujian Balanced Dataset	66
Tabel 4. 23 Confusion Matrix Kernel Linear Pengujian Balanced Dataset	66
Tabel 4. 24 Hasil Akurasi Perbandingan Pengujian Unbalanced dan Balanced	67
Tabel 4. 25 Hasil Confusion Matrix Pengujian dengan Pelabelan Lexicon Based	68
Tabel 4. 26 Hasil Confusion Matrix Pengujian dengan tanpa Pelabelan Manual	68
Tabel 4. 27 Hasil Confusion Matrix pada Uji Silang	69
Tabel 4. 28 Hasil Akurasi Perbandingan Pengujian	69

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi SVM[51]	26
Gambar 3. 1 Diagram Alir Penelitian	31
Gambar 3. 2 Pengumpulan Data dengan Octoparse	32
Gambar 3. 3 Pengolahan Data dengan Pelabelan Manual	36
Gambar 3. 4 Pengolahan Data dengan Pelabelan Lexicon Based	38
Gambar 3. 5 Alir Diagram Preprocessing	39
Gambar 3. 6 Alir Pembuatan Kamus Lexicon	40
Gambar 3. 7 Alir Pelabelan Lexicon	42
Gambar 4. 1 Distribusi Nilai Bobot Dataset Hasil Pelabelan	51
Gambar 4. 2 Hasil Pengujian SVM	62
Gambar 4. 3 Grafik Hasil Klasifikasi Label Prediktif dan Aktual	64

DAFTAR SINGKATAN

BP Tapera	:	Badan Pengelolaan Tabungan Pengelolaan Rakyat
FN	:	<i>False Negative</i>
FNe	:	<i>False Neutral</i>
FNt	:	<i>False Netral</i>
FP	:	<i>False Positive</i>
KNN	:	<i>K-Nearest Neighbour</i>
MMH	:	<i>Maximum Margin Hyperplane</i>
PP	:	Peraturan Pemerintah
RBF	:	<i>Radial Basis Function</i>
SRM	:	<i>Structural Risk Minimization</i>
SVM	:	<i>Support Vector Machine</i>
Tapera	:	Tabungan Pengelolaan Rakyat
TF-IDF	:	<i>Term Frequency-Inverse Document Frequency</i>
TN	:	<i>True Negative</i>
TNt	:	<i>True Negative</i>
TNe	:	<i>True Neutral</i>
TP	:	<i>True Positive</i>
UU	:	Undang-Undang

DAFTAR LAMPIRAN

Lampiran 1. Perhitungan Hasil Confusion Matrix Tiap Kernel	79
Lampiran 2. Source Code Pengujian	81
Lampiran 3. Source Code Perbandingan Pengujian	90