

TUGAS AKHIR

**ANALISIS SENTIMEN MASYARAKAT TERHADAP
TRAGEDI KANJURUHAN DENGAN METODE
*SUPPORT VECTOR MACHINE***



MARTIN PARHUSIP
19102042

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO
2023**

TUGAS AKHIR

**ANALISIS SENTIMEN MASYARAKAT TERHADAP
TRAGEDI KANJURUHAN DENGAN METODE
*SUPPORT VECTOR MACHINE***

***SENTIMENT ANALYSIS OF THE PUBLIC TOWARDS
THE KANJURUHAN TRAGEDY WITH THE SUPPORT
VECTOR MACHINE METHOD***

Disusun Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer



MARTIN PARHUSIP
19102042

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO
2023**

HALAMAN PERSETUJUAN PEMBIMBING

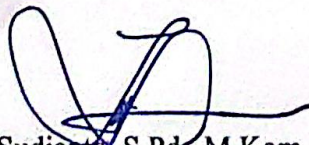
**ANALISIS SENTIMEN MASYARAKAT TERHADAP
TRAGEDI KANJURUHAN DENGAN METODE
*SUPPORT VECTOR MACHINE***

***SENTIMENT ANALYSIS OF THE PUBLIC TOWARDS
THE KANJURUHAN TRAGEDY WITH THE SUPPORT
VECTOR MACHINE METHOD***

Dipersiapkan dan Disusun oleh
Martin Parhusip
19102042

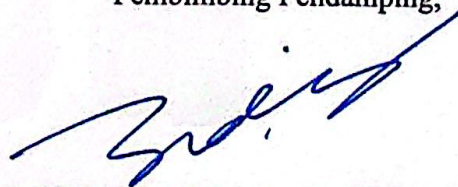
**Fakultas Informatika
Institut Teknologi Telkom Purwokerto
Pada Tanggal: 6 Juli 2023**

Pembimbing Utama,



Sudianto, S.Pd., M.Kom.
NIDN 0605049301

Pembimbing Pendamping,



Dr.H.Tri Ginanjar Laksana, M.Cs., M.Kom.
NIDN 0407088502

HALAMAN PENGESAHAN

**ANALISIS SENTIMEN MASYARAKAT TERHADAP
TRAGEDI KANJURUHAN DENGAN METODE
*SUPPORT VECTOR MACHINE***

***SENTIMENT ANALYSIS OF THE PUBLIC TOWARDS
THE KANJURUHAN TRAGEDY WITH THE SUPPORT
VECTOR MACHINE METHOD***

Dipersiapkan dan Disusun oleh
Martin Parhusip
19102042

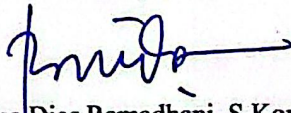
Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tugas Akhir
Pada Selasa, 17 Juli 2022.

Penguji I,



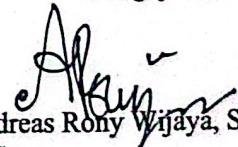
Amalia Beladina Arifa,
S.Pd., M.Cs.
NIDN 0606019201

Penguji II,



Rima Dias Ramadhani, S.Kom.,
M.Kom
NIDN 0517108101

Penguji III,



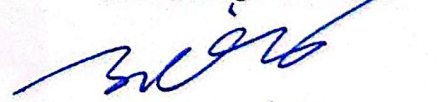
Andreas Rony Wijaya, S.Mat.,
M.Sc
NIDN 0630059602

Pembimbing Utama,



Sudianto, S.Pd., M.Kom
NIDN 0605049301

Pembimbing Pendamping



Dr.H.Tri Ginanjar Laksana,
M.Cs., M.Kom.
NIDN 0407088502

Dekan Fakultas Informatika,



Auliya Burhanuddin, S.Si., M.Kom.
NIK 19820008

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Martin Parhusip**
NIM : **19102042**
Program Studi : **Teknik Informatika**

Menyatakan bahwa Tugas Akhir dengan judul berikut:

**ANALISIS SENTIMEN MASYARAKAT TERHADAP TRAGEDI
KANJURUHAN DENGAN METODE *SUPPORT VECTOR MACHINE***

Dosen Pembimbing Utama : **Sudianto, S.Pd., M.Kom**
Dosen Pembimbing Pendamping : **Dr. H. Tri Laksana, M.Cs., M.Kom**

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 tanggungjawab 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.

Purwokerto, 2 November 2022,


(Martin Parhusip)

KATA PENGANTAR

Pertama penulis panjatkan puji dan syukur kepada kehadiran Tuhan Yang Maha Esa atas rahmat dan berkatNya sehingga penulis bisa mengerjakan tugas akhir yang berjudul “Analisis Sentimen Masyarakat terhadap Tragedi Kanjuruhan dengan Metode *Support Vector Machine*” hingga selesai tepat waktu. Penulis dalam kesempatan ini mengucapkan terima kasih kepada:

1. Orang tua penulis yang mendukung, mendoakan, serta memberikan banyak bantuan untuk penulis selama pengerjaan tugas akhir.
2. Ibu Dr. Tenia Wahyuningrum, S.Kom., M.T selaku Rektor Institut Teknologi Telkom Purwokerto.
3. Bapak Auliya Burhanuddin, S.Si., M.Kom. selaku Dekan Fakultas Informatika.
4. Ibu Amalia Beladonna Arifa, S.Pd., M.Cs selaku Kepala Program Studi S1 Teknik Informatika Institut Teknologi Telkom Purwokerto.
5. Bapak Sudianto, S.Pd., M.Kom. selaku dosen pembimbing utama dan Bapak Dr. H. Tri Ginanjar Laksana, S. Kom., MCs, M. Kom selaku dosen pembimbing pendamping yang memberikan bimbingan serta arahan hingga tugas akhir penulis ini dapat selesai dengan baik dan tepat waktu.
6. Keluarga yang telah memberikan dukungan dan doa untuk penulis ketika pengerjaan tugas akhir.
7. Teman-teman mahasiswa Institut Teknologi Telkom Purwokerto dan seluruh pihak yang tidak bisa penulis sebutkan satu persatu.

Penulis menyadari bahwa tugas akhir ini belum sempurna dan mengharapkan kritik serta saran yang membangun dari pembaca, dosen, dan rekan mahasiswa. Penulis berharap tugas akhir ini dapat bermanfaat, meningkatkan wawasan pembaca, dan menginspirasi penelitian lebih lanjut di bidang yang sama.

Purwokerto, 21 Januari 2023



Martin Parhusip

DAFTAR ISI

HALAMAN PERSETUJUAN PEMBIMBING.....	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....	iii
KATA PENGANTAR.....	iv
DAFTAR ISI	v
DAFTAR GAMBAR.....	viii
DAFTAR TABEL	ix
DAFTAR SINGKATAN	xi
DAFTAR ISTILAH.....	xii
DAFTAR LAMPIRAN	xiv
ABSTRAK.....	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah	1
1.2 Pertanyaan Penelitian	3
1.3 Perumusan Masalah.....	3
1.4 Batasan Masalah	4
1.5 Tujuan Penelitian.....	4
1.6 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	6
2.1. Tinjauan Pustaka	6
2.2. Dasar Teori	16
2.2.1 Analisis Sentimen	16
2.2.2 <i>Crawling</i>	16
2.2.3 <i>Preprocessing</i>	17
2.2.4 <i>Wordcloud</i>	17
2.2.5 Pelabelan	17

2.2.6	<i>Indonesian RoBERTa Sentiment Classifier Inference</i>	18
2.2.7	Pembobotan.....	18
2.2.8	<i>Support Vector Machine</i>	19
2.2.9	Parameter Evaluasi.....	22
2.2.10	<i>Twitter</i>	24
2.2.11	<i>Youtube</i>	24
BAB III METODOLOGI PENELITIAN		25
3.1	Objek dan Subjek Penelitian	25
3.1.1	Subjek Penelitian	25
3.1.2	Objek Penelitian.....	25
3.2	Alat dan Bahan Penelitian	25
3.2.1	Alat Penelitian	25
3.2.2	Bahan Penelitian	26
3.3	Diagram Alir Penelittian.....	26
3.3.1	Identifikasi dan Perumusan masalah	27
3.3.2	Studi Literatur	27
3.3.3	Pengumpulan Data.....	28
3.3.4	<i>Preprocessing</i>	29
3.3.5	Pembuatan <i>Wordcloud</i>	31
3.3.6	Pelabelan Data	32
3.3.7	Klasifikasi dengan <i>SVM</i>	34
3.3.8	Evaluasi dan Kesimpulan	36
BAB IV HASIL DAN PEMBAHASAN		37
4.1	Hasil Pengujian.....	37
4.1.1	Pengumpulan Data	37
4.1.2	<i>Preprocessing</i> Data.....	38
4.1.3	<i>Labeling</i> Data.....	42
4.1.4	<i>Split</i> Data.....	44
4.1.5	Pembobotan Data	45
4.1.6	Implementasi <i>SVM</i>	47

4.3	Evaluasi dan Pembahasan.....	52
BAB V	KESIMPULAN DAN SARAN	62
DAFTAR	PUSTAKA	64
LAMPIRAN	68

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Dataset <i>Linear</i> [27].....	20
Gambar 2. 2 Ilustrasi Dataset <i>non-Linear</i> [27]	21
Gambar 3. 1 Diagram Alir Penelitian	27
Gambar 3. 2 Diagram Alir Pengumpulan Data.....	28
Gambar 3. 3 Diagram Alir <i>Preprocessing</i>	29
Gambar 3. 4 <i>Wordcloud</i> dari <i>tweet</i> tragedi kanjuruhan	32
Gambar 3. 5 <i>Wordcloud</i> dari komen <i>youtube</i> Tragedi.....	32
Gambar 3. 6 Diagram Alir Pelabelan Data	33
Gambar 3. 7 Diagram Alir Klasifikasi	34
Gambar 4. 1 Sepuluh hashtag terbanyak pada <i>tweet</i> tragedi kanjuruhan.....	39
Gambar 4. 2 Diagram batang distribusi sentimen dengan 3 label.....	43
Gambar 4. 3 Diagram batang distribusi sentimen dengan 2 label.....	43
Gambar 4. 4 Hasil uji coba kernel <i>SVM</i> pada 3 label sentimen.	58
Gambar 4. 5 Hasil uji coba kernel <i>SVM</i> pada 2 label sentimen.	59

DAFTAR TABEL

Tabel 2. 1 <i>Literature review</i> dari penelitian sebelumnya.....	9
Tabel 2. 2 Rumus kernel <i>SVM</i>	21
Tabel 3. 1 Alat Penelitian.....	25
Tabel 3. 2 Contoh <i>Tweet</i> /komentar tentang tragedi kanjuruhan	29
Tabel 3. 3 Contoh hasil <i>preprocessing</i> data	30
Tabel 3. 4 Contoh <i>dataset</i> yang sudah dilabeli	33
Tabel 3. 5 Contoh pembobotan dengan <i>TF-IDF</i>	35
Tabel 3. 6 Parameter kernel <i>SVM</i> yang digunakan	35
Tabel 4. 1 Contoh Hasil Pengambilan Data <i>Tweet</i> dan <i>Youtube</i>	37
Tabel 4. 2 Contoh hasil dari <i>Cleaning</i>	39
Tabel 4. 3 Contoh hasil dari <i>Case Folding</i>	40
Tabel 4. 4 Jumlah teks setelah penyaringan.....	40
Tabel 4. 5 Contoh hasil normalisasi teks	41
Tabel 4. 6 Contoh hasil dari <i>Tokenizing</i>	41
Tabel 4. 7 Contoh hasil dari penghapusan <i>Stopwords</i>	42
Tabel 4. 8 Contoh hasil dari <i>Lematization</i>	42
Tabel 4. 9 Hasil proses <i>labeling</i>	44
Tabel 4. 10 <i>Splitting</i> data untuk 3 sentimen.....	45
Tabel 4. 11 <i>Splitting</i> data untuk 2 sentimen.....	45
Tabel 4. 12 Hasil pembobotan <i>TF-IDF</i> pada data latih	46
Tabel 4. 13 Hasil pembobotan <i>TF-IDF</i> pada data uji	47
Tabel 4. 14 Nilai <i>Hyperparameter</i>	48
Tabel 4. 15 Nilai <i>Hyperparameter</i> terbaik.....	48
Tabel 4. 16 Sample data teks tragedi Kanjuruhan.....	49
Tabel 4. 17 Contoh data uji	51
Tabel 4. 18 <i>Confusion Matrix</i> 3x3	52
Tabel 4. 19 <i>Confusion Matrix</i> 2x2	53
Tabel 4. 20 <i>Confusion matrix</i> untuk kernel <i>linear</i> dengan 3 label sentimen	55
Tabel 4. 21 <i>Confusion matrix</i> untuk kernel <i>polinomial</i> dengan 3 label sentimen.	55

Tabel 4. 22	<i>Confusion matrix</i> untuk kernel <i>sigmoid</i> dengan 3 label sentimen	56
Tabel 4. 23	<i>Confusion matrix</i> untuk kernel <i>RBF</i> dengan 3 label sentimen.....	56
Tabel 4. 24	<i>Confusion matrix</i> untuk kernel <i>linear</i> dengan 2 label sentimen	56
Tabel 4. 25	<i>Confusion matrix</i> untuk kernel <i>polynomial</i> dengan 2 label sentimen	57
Tabel 4. 26	<i>Confusion matrix</i> untuk kernel <i>sigmoid</i> dengan 2 label sentimen	57
Tabel 4. 27	<i>Confusion matrix</i> untuk kernel <i>RBF</i> dengan 2 label sentimen.....	57
Tabel 4. 28	Hasil uji fungsi kernel pada 3 label sentimen	58
Tabel 4. 29	Hasil uji fungsi kernel pada 2 label sentimen	58
Tabel 4. 30	Perbandingan antara data aktual dan prediksi untuk 3 label	60
Tabel 4. 31	Perbandingan antara data aktual dan prediksi untuk 2 label.....	60

DAFTAR SINGKATAN

<i>API</i>	: <i>Application Programming Interface</i>
<i>CNN</i>	: <i>Convolutional Neural Network</i>
<i>CSV</i>	: <i>Comma Separated Value</i>
<i>FN</i>	: <i>False Negative</i>
<i>FP</i>	: <i>False Positive</i>
<i>KNN</i>	: <i>K Nearest Neighbor</i>
<i>NB</i>	: <i>Naïve Bayes</i>
<i>RBF</i>	: <i>Radial Basic Function</i>
<i>SMOTE</i>	: <i>Synthetic Minority Over-sampling Technique</i>
<i>SVM</i>	: <i>Support Vector Machine</i>
<i>TF-IDF</i>	: <i>Term Frequency – Invers Document Frequency</i>
<i>TN</i>	: <i>True Negative</i>
<i>TP</i>	: <i>True Positive</i>

DAFTAR ISTILAH

- Casefolding* : Proses mengubah semua huruf dalam teks menjadi huruf kecil untuk mengurangi variasi.
- Confusion Matrix* : Matriks yang menunjukkan hasil prediksi klasifikasi yang dibandingkan dengan label kelas sebenarnya.
- Deep Learning* : Cabang dari *machine learning* yang fokus pada penggunaan jaringan saraf tiruan dengan banyak lapisan untuk mempelajari representasi data yang kompleks.
- Grid Search* : Metode untuk mencari kombinasi terbaik dari *hyperparameter* dalam algoritma *machine learning*.
- Hyperparameter* : Parameter yang tidak diatur secara otomatis oleh algoritma machine learning dan harus diatur secara manual sebelum pelatihan model.
- Hyperplane* : Dalam konteks *machine learning*, ini adalah batas keputusan yang memisahkan data dengan label kelas yang berbeda dalam ruang fitur.
- Kernel Tricks* : Teknik yang digunakan dalam algoritma *SVM* untuk memetakan data ke ruang fitur yang lebih tinggi agar dapat dipisahkan secara *linear*.
- Lemmatization* : Mengubah kata menjadi bentuk dasarnya berdasarkan konteks dan maknanya
- Machine Learning* : Cabang dari kecerdasan buatan yang melibatkan pembelajaran komputer untuk mengidentifikasi pola dan membuat prediksi berdasarkan data input
- Normalization* : Proses mengubah data menjadi rentang yang lebih umum atau normal, seperti mengubah skala atau menghilangkan perbedaan satuan. *Oversampling*:

- Teknik untuk menambahkan kelas minoritas dalam data latih dengan membuat salinan atau sintesis data baru.
- Stemming* : Proses menghapus akhiran atau awalan kata untuk menghasilkan kata dasar atau bentuk kata yang lebih sederhana.
- Stopwords* : Kata-kata umum yang dianggap tidak memiliki nilai prediktif dan dihapus dari teks saat pemrosesan bahasa alami.
- Supervised Learning* : Pendekatan dalam *machine learning* di mana model dilatih menggunakan data yang berlabel, dengan tujuan untuk membuat prediksi pada data yang belum terlihat sebelumnya.
- Tuning* : Proses mengoptimalkan *hyperparameter* atau konfigurasi model *machine learning* untuk mencapai performa terbaik.
- Undersampling* : Teknik untuk mengurangi jumlah sampel dalam kelas mayoritas dalam data latih untuk menjaga keseimbangan dengan kelas minoritas.

DAFTAR LAMPIRAN

Lampiran 1 Data hasil <i>crawling</i> dari <i>Twiter</i>	68
Lampiran 2 Data hasil <i>crawling</i> dari <i>Youtube</i>	68
Lampiran 3 <i>Source code</i> untuk <i>Preprocessing</i>	69
Lampiran 4 <i>Source code</i> untuk <i>Labeling</i>	70
Lampiran 5 <i>Source code</i> Klasifikasi dengan <i>SVM</i> 2 label	71
Lampiran 6 <i>Source code</i> klasifikasi dengan <i>SVM</i> 3 label	73