

**PERBANDINGAN PERFORMA TF-IDF DAN WORD2VEC  
UNTUK ANALISIS SENTIMEN *CYBERBULLYING*  
MENGUNAKAN METODE SUPPORT VECTOR  
MACHINE (SVM)**

**SKRIPSI**



Oleh :

**AHMAD HILMAN DANI**

**20081010057**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2024**

**LEMBAR PENGESAHAN  
SKRIPSI**

**Judul : PERBANDINGAN PERFORMA TF-IDF DAN WORD2VEC  
UNTUK ANALISIS SENTIMEN *CYBERBULLYING*  
MENGUNAKAN METODE SUPORT VECTOR MACHINE  
(SVM)**

**Oleh : Ahmad Hilman Dani**

**NPM : 20081010057**

**Telah Diseminarkan Dalam Ujian Skripsi Pada :**

**Hari Selasa, Tanggal 21 Mei 2024**

**Mengetahui**

**Dosen Pembimbing**

1.

**Eva Yulia P, S.Kom., M.Kom**

**NIP: 19890705 2021212 002**

2.

**Retno Mumpuni, S.Kom., M.Sc  
M.Kom**

**NPT : 172198 70 716054**

**Dosen Penguji**

1.

**Yisti Vita Via, S.ST. M.Kom**

**NIP: 19860425 2021212 001**

2.

**Agung Mustika Rizki, S.Kom.,**

**NIP: 19930725 202203 1008**

**Menyetujui**



**Prof. Dr. Ir. Novirina Hendrasarie, MT.**

**NIP : 19681126 199403 2 001**

**Koordinator Program Studi  
Teknik Informatika**

**Fetty Tri Anggraeny, S.Kom.  
M.Kom**

**NIP: 19820211 2021212 005**

## SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Ahmad Hilman Dani

NPM : 20081010057

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

“PERBANDINGAN PERFORMA TF-IDF DAN WORD2VEC UNTUK ANALISIS SENTIMEN CYBERBULLYING MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)”

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dari juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.



Surabaya, 30 Mei 2024



Ahmad Hilman Dani  
NPM. 20081010057

# PERBANDINGAN PERFORMA TF-IDF DAN WORD2VEC UNTUK ANALISIS SENTIMEN CYBERBULLYING MENGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

**Nama Mahasiswa : Ahmad Hilman Dani**

**NPM : 20081010057**

**Program Studi : Teknik Informatika**

**Dosen Pembimbing : Eva Yulia Puspaningrum, S.Kom., M.Kom**

**Retno Mumpuni, S.Kom., M.Sc**

## **Abstrak**

*Cyberbullying* merupakan satu diantara beberapa dampak negatif dari media sosial yang semakin marak terjadi, sehingga, dilakukan analisis sentimen terhadap konten *cyberbullying* dengan harapan dapat mendeteksi dan menyaring dan konten yang terindikasi *bullying* di media sosial. Analisis sentimen memerlukan sebuah metode pembobotan kata untuk mengkonversi kata menjadi angka karena komputer tidak mampu memproses kata secara langsung. Maka dari itu, metode pembobotan kata yang digunakan menjadi sangat penting karena dapat berdampak langsung terhadap proses analisis sentimen. Berdasarkan hal ini, peneliti melakukan penelitian yang membandingkan performa dua teknik pembobotan kata, yakni Term Frequency-Inverse Document Frequency (TF-IDF) dan Word2Vec untuk analisis sentimen *cyberbullying* menggunakan metode Support Vector Machine (SVM).

TF-IDF merupakan metode pembobotan kata yang sederhana dengan cara mengkonversi teks menjadi vektor numerik berdasarkan frekuensi kemunculan kata dalam sebuah dokumen dan keseluruhan korpus tanpa memperhatikan hubungan semantik antar kata, padahal hubungan semantik antar kata merupakan hal yang tidak dapat diabaikan. Adapun, Word2Vec merupakan metode pembobotan kata yang dapat menangkap hubungan semantik antar kata. Maka dari itu, secara teori, seharusnya TF-IDF memiliki performa yang tidak lebih baik daripada Word2Vec untuk analisis sentimen.

***Kata kunci:*** Support Vector Machine, TF-IDF, Word2Vec, Sentimen Analisis, Cyberbullying.

## **KATA PENGANTAR**

Alhamdulillah, atas izin Allah Azza Wa Jalla semata yang membuat penulis dapat menyelesaikan penelitian dan laporan yang berjudul:

### **“PERBANDINGAN PERFORMA TF-IDF DAN WORD2VEC UNTUK ANALISIS SENTIMEN CYBERBULLYING MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)”**

Penulis mengucapkan dengan hati yang sangat ikhlas dan sungguh-sungguh bahwa selesainya laporan ini bukan karena kepiwaian penulisan penulis, namun semata-mata karena rahmat Allah Azza Wa Jalla semata, maka dari itu penulis sangat bersyukur kepada-Nya dan Dia Azza Wa Jalla telah memerintahkan hambanya untuk mengungkapkan rasa syukur tersebut melalui lisan dengan berkata “Alhamdulillah”, maka dari itu, dengan hati yang sangat berbahagia, ikhlas, dan bersungguh-sungguh, penulis mengucapkan, Alhamdulillah.

Seluruh kekurangan yang ada di laporan ini merupakan kesalahan penulis dan penulis sadar akan hal itu. Sehingga, penulis menerima kritik dan saran yang membangun untuk memperbaiki laporan ini

Surabaya, 30 Mei 2024

Penulis

## UCAPAN TERIMA KASIH

Dalam penyelesaian laporan ini, Alhamdulillah, Allah Azza Wa Jalla telah menjadikan beberapa pihak sebagai peranta untuk menolong penulis dalam menyelesaikan laporan ini.

Maka dengan ini, penulis ingin mengucapkan terima kasih kepada:

1. Ibu Anitawati dan Bapak Pram Wedi Nugraha yang telah menjadi orang yang tidak pernah menyerah kepada anaknya, selalu mendukung, menyemangati, dan menenangkan pikiran anaknya sehingga atas izin Allah Azza Wa Jalla, putra mereka berdua berhasil menyelesaikan laporan ini
2. Bapak Prof. Dr. Ir. Akhmad Fauzi, MMT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Dr. Novirina Hendrasarie, S.T, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom. selaku Koordinator Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom selaku dosen pembimbing pertama yang benar-benar telah banyak membantu penulis selama pengerjaan dari awal hingga selesai.
6. Ibu Retno Mumpuni, S.Kom., M.Sc selaku dosen pembimbing kedua yang benar-benar telah banyak membantu penulis selama pengerjaan dari awal hingga selesai.
7. Seluruh Dosen Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan ilmu yang berguna dan bermanfaat selama perkuliahan.
8. Seluruh pihak yang penulis tidak dapat sebutkan seluruhnya.

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	ii
KATA PENGANTAR .....	vi
UCAPAN TERIMA KASIH.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian .....	4
1.4 Manfaat Penelitian .....	5
1.5 Batasan Masalah .....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu.....	6
2.2 Dasar Teori .....	9
2.2.1 <i>Bullying</i> .....	9
2.2.2 Analisis sentimen .....	9
2.2.1 Web Scrapping .....	9
2.2.2 Case Folding .....	9
2.2.3 Cleaning .....	10
2.2.4 Tokenization.....	10
2.2.5 Stopword Removal.....	10
2.2.6 Stemming .....	11
2.2.7 TF-IDF .....	11



2.2.8	Word2Vec .....	11
2.2.9	Support Vector Machine .....	17
2.2.10	Confusion Matrix .....	20
BAB III METODOLOGI.....		23
3.1	Tahapan Pengerjaan Penelitian .....	23
3.2	Studi literatur .....	23
3.3	Desain flowchart untuk menerapkan model sentiment analisis .....	23
3.4	Akuisisi data .....	25
3.6	Pembobotan Kata .....	31
3.6.1	TF-IDF.....	31
3.6.2	Word2Vec CBOW.....	34
3.6.3	Word2Vec Skip-gram.....	38
3.7	Support Vector Machine (SVM).....	41
3.8	Pengujian model dengan Confusion Matrix.....	42
BAB IV HASIL DAN PEMBAHASAN .....		44
4.1	Proses Akuisisi Data .....	44
4.2	Preprocessing Data.....	45
4.3	Pembagian Data Training dan Testing .....	51
4.4	Pembobotan Kata .....	52
4.4.1	TF-IDF.....	52
4.4.2	Word2Vec .....	53
4.5	Proses dan Hasil Klasifikasi Menggunakan SVM.....	56
4.5.1	Skenario 1: TF-IDF dengan SVM karnel linear .....	57
4.5.2	Skenario 2: TF-IDF dengan SVM karnel RBF .....	58
4.5.3	Skenario 3: TF-IDF dengan SVM karnel polynomial .....	60
4.5.4	Skenario 4: Word2Vec CBOW dengan SVM karnel linear .....	62

4.5.5	Skenario 5: Word2Vec CBOW dengan SVM karnel RBF .....	64
4.5.6	Skenario 6: Word2Vec CBOW dengan SVM karnel <i>polynomial</i> 66	
4.5.7	Skenario 7: Word2Vec Skip-gram dengan SVM karnel linear...	68
4.5.8	Skenario 8: Word2Vec Skip-gram dengan SVM karnel RBF ....	70
4.5.9	Skenario 9: Word2Vec Skip-gram dengan SVM karnel <i>polynomial</i> .....	72
4.6	Analisis hasil skenario uji.....	74
BAB V PENUTUP.....		80
5.1	Kesimpulan.....	80
5.2	Saran	80
LAMPIRAN SUMBER KODE .....		84

## DAFTAR TABEL

Tabel 2. 1 Contoh Case Folding.....	9
Tabel 2. 2 Contoh cleaning .....	10
Tabel 2. 3 Contoh tokenization .....	10
Tabel 2. 4 Contoh stopword removal.....	10
Tabel 2. 5 Contoh stemming .....	11
Tabel 2. 6 Tabel Confusion Matrix .....	20
Tabel 3. 1 Contoh dokumen yang ada di korpus.....	32
Tabel 3. 2 Perhitungan TF tiap kata dalam dokumen .....	32
Tabel 3. 3 Perhitungan IDF tiap kata .....	33
Tabel 3. 4 Hasil akhir TF-IDF.....	33
Tabel 3. 5 One-hot-encoding.....	35
Tabel 3. 6 One-hot-encoding.....	38
Tabel 4. 1 Hasil banyaknya data uji yang diprediksi sebagai bullying atau non-bullying dengan metrik TP, FP, TN, dan FN .....	75
Tabel 4. 2 Nilai akurasi, presisi, recall, dan f1-score masing-masing skenario uji	76

## DAFTAR GAMBAR

Gambar 2. 1 ilustrasi CBOW .....	12
Gambar 2. 2 Ilustrasi Skip-gram .....	16
Gambar 2. 3 ilustrasi SVM .....	18
Gambar 3. 1 Tahapan pengerjaan penelitian.....	23
Gambar 3. 2 Flowchart penerapan model .....	23
Gambar 3. 3 Flowchart akuisis data dan labeling .....	25
Gambar 3. 4 Isi file hasil unduhan di kaggle.com.....	26
Gambar 3. 5 Flowchart preprocessing data.....	26
Gambar 3. 6 Contoh hasil pelabelan data.....	27
Gambar 3. 7 Contoh sentimen setelah dilakukan case folding .....	27
Gambar 3. 8 Contoh sentimen setelah dilakukan case folding dan cleaning .....	28
Gambar 3. 9 Contoh sentimen setelah dilakukan case folding, cleaning, dan remove repeating character .....	28
Gambar 3. 10 Contoh sentimen setelah dilakukan case folding, cleaning, remove repeating character, dan stemming.....	29
Gambar 3. 11 Contoh sentimen setelah dilakukan case folding, cleaning, remove repeating character, stemming, dan stopwrod removal.....	29
Gambar 3. 12 Gambar tabel daftar kata stopwords di sastrawi.....	30
Gambar 3. 13 Contoh sentimen setelah dilakukan case folding, cleaning, remove repeating character, stemming, stopword removal, dan tokenization.	30
Gambar 3. 14 Flowchart TF-IDF.....	31
Gambar 3. 15 Ilustrasi model CBOW .....	34
Gambar 3. 16 Ilustrasi model skip-gram.....	38
Gambar 3. 17 flowchart SVM.....	41
Gambar 4. 1 Tampilan kaggle.com .....	44
Gambar 4. 2 File .zip hasil dari download data di kaggle.com.....	44
Gambar 4. 3 Wujud dari data yang didapatkan penulis dari kaggle.com .....	45
Gambar 4. 4 Wujud data setelah dihapus kolom yang tidak diperlukan.....	47
Gambar 4. 5 Wujud data setelah dihapus kolom yang tidak diperlukan.....	47
Gambar 4. 6 Sentimen setelah dilakukan Case folding .....	48
Gambar 4. 7 Sentimen setelah dilakukan case folding dan cleaning .....	49

Gambar 4. 8 data setelah dilakukan case folding, cleaning, dan stemming.....	49
Gambar 4. 9 Contoh sentimen setelah dilakukan case folding, cleaning, stemming, dan stopwrod removal .....	50
Gambar 4. 10 Contoh sentimen setelah dilakukan case folding, cleaning, stemming, dan stopwrod removal, tokenization .....	50
Gambar 4. 11 Hasil akhir data setelah melewati preprocessing data.....	51
Gambar 4. 12 Contoh hasil TF-IDF .....	52
Gambar 4. 13 Contoh vektor dari kata “uang” .....	53
Gambar 4. 14 Tingkat keterkaitan antara kata “uang” dan kata “duit” .....	54
Gambar 4. 15 Confusion Matrix Tabel SVM TF-IDF linear .....	57
Gambar 4. 16 Hasil akurasi, presisi, recall, f1-scoreTF-IDF SVM linear .....	58
Gambar 4. 17 Confusion matrix tabel TF-IDF SVM RBF .....	59
Gambar 4. 18 Hasil akurasi, presisi, recall, f1-score TF-IDF SVM RBF.....	60
Gambar 4. 19 Confusion Matrix Tabel TF-IDF SVM polynomial.....	61
Gambar 4. 20 Hasil akurasi, presisi, recall, f1-score TF-IDF SVM Polynomial...	62
Gambar 4. 21 Confusion Matrix Tabel Word2Vec CBOW SVM karnel linear .....	63
Gambar 4. 22 Hasil akurasi, presisi, recall, f1-score Word2Vec CBOW SVM linear .....	64
Gambar 4. 23 Confusion Matrix Tabel Word2Vec CBOW SVM karnel RBF .....	65
Gambar 4. 24 Hasil akurasi, presisi, recall, f1-score Word2Vec CBOW SVM RBF .....	66
Gambar 4. 25 Confusion Matrix Tabel Word2Vec CBOW SVM karnel polynomial .....	67
Gambar 4. 26 Hasil akurasi, presisi, recall, f1-score Word2Vec CBOW SVM polynomial .....	68
Gambar 4. 27 Confusion Matrix Tabel Word2Vec Skip-gram SVM karnel linear .....	69
Gambar 4. 28 Hasil akurasi, presisi, recall, f1-score Word2Vec Skip-gram SVM linear.....	70
Gambar 4. 29 Confusion Matrix Tabel Word2Vec Skip-gram SVM karnel RBF	71

Gambar 4. 30 Hasil akurasi, presisi, recall, f1-score Word2Vec Skip-gram SVM RBF .....	72
Gambar 4. 31 Confusion Matrix Tabel Word2Vec Skip-gram SVM karnel polynomial .....	73
Gambar 4. 32 Hasil akurasi, presisi, recall, f1-score Word2Vec Skip-gram SVM polynomial .....	74