

**DETEKSI UNSUR DEFAMASI  
PADA MEDIA SOSIAL *TWITTER* MENGGUNAKAN  
KLASIFIKASI *SUPPORT VECTOR MACHINE***

**Skripsi**

**Disusun untuk memenuhi salah satu syarat  
memperoleh gelar Sarjana Komputer**



**Disusun oleh:**

**Candra Febrian Tri Pambudi  
3145162479**




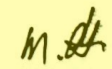


**PROGRAM STUDI ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI JAKARTA  
2023**

# LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

## DETEKSI UNSUR DEFAMASI PADA MEDIA SOSIAL TWITTER MENGGUNAKAN KLASIFIKASI SUPPORT VECTOR MACHINE

Nama : Candra Febrian Tri Pambudi

No. Registrasi : 3145162479

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: <u>Prof. Dr. Muktiningsih N. M.Si.</u> NIP. 196405111989032001		31/8/2023
Wakil Penanggung Jawab			
Wakil Dekan I	: <u>Dr. Esmar Budi, S.Si., MT.</u> NIP. 197207281999031002		31/8/23
Ketua	: <u>Med Irzal, M.Kom.</u> NIP. 197706152003121001		24/08/2023
Sekretaris	: <u>Ari Hendarno, S.Pd., M.Kom.</u> NIP. 198811022022031002		24/08/2023
Penguji	: <u>Muhammad Eka Suryana, M.Kom.</u> NIP. 198512232012121002		25/08/2023
Pembimbing I	: <u>Dr. Ria Arafyah, M.Si.</u> NIP. 197511212005012004		24/08/2023
Pembimbing II	: <u>Drs. Mulyono, M.Kom.</u> NIP. 196605171994031003		24/08/2023

Dinyatakan lulus ujian skripsi tanggal: 21 Agustus 2023

## LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul Deteksi Unsur Defamasi Pada Media Sosial Twitter Menggunakan Klasifikasi *Support Vector Machin* yang disusun sebagai syarat untuk memperoleh gelar Sarjana komputer dari Program Studi Ilmu Komputer Universitas Negeri Jakarta adalah karya ilmiah saya dengan arahan dari dosen pembimbing.

Sumber informasi yang diperoleh dari penulis lain yang telah dipublikasikan yang disebutkan dalam teks skripsi ini, telah dicantumkan dalam Daftar Pustaka sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Jika dikemudian hari ditemukan sebagian besar skripsi ini bukan hasil karya saya sendiri dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sanding dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Jakarta, 30 Juli 2023



Candra Febrian Tri Pambudi  
NRM. 3145162479



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN  
UNIVERSITAS NEGERI JAKARTA  
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220  
Telepon/Faksimili: 021-4894221  
Laman: [lib.unj.ac.id](http://lib.unj.ac.id)

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Candra Febrian Tri Pambudi  
NIM : 3145162479  
Fakultas/Prodi : FMIPA/Ilmu Komputer  
Alamat email : [appletjandra@gmail.com](mailto:appletjandra@gmail.com)

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi     Tesis     Disertasi     Lain-lain (.....)

yang berjudul :

**DETEKSI UNSUR DEFAMASI PADA MEDIA SOSIAL TWITTER  
MENGUNAKAN KLASIFIKASI SUPPORT VECTOR MACHINE**

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta Selasa 25 September 2023

Penulis

( **Candra Febrian Tri P.** )  
*nama dan tanda tangan*

## ABSTRAK

**CANDRA FEBRIAN TRI PAMBUDI. Deteksi Unsur Defamasi Pada Media Sosial Twitter Menggunakan Klasifikasi *Support Vector Machine*.** Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. Agustus 2023. Di bawah bimbingan Ibu Dr. Ria Arafiyah, M.Si. dan Bapak Drs. Mulyono, M.Kom.

Defamasi atau Pencemaran nama baik adalah perbuatan yang secara sengaja membuat nama baik atau reputasi orang tersebut menjadi tercemar atau buruk. Hal ini menimbulkan pembunuhan karakter atau *character assassination* yang merupakan pelanggaran HAM. Skripsi ini bertujuan untuk membangun model deteksi defamasi dengan klasifikasi SVM pada data Twitter dan SafeNet, menghitung performa SVM dan mengetahui pengaruh TF-IDF serta SVD pada ekstraksi fitur terhadap performa model. Berdasarkan SKB 3 Menteri poin g, tindakan yang dikatakan pencemaran nama baik yaitu tindakan yang mengandung unsur penuduhan. Sedangkan yang bukan termasuk defamasi, narasi yang mengandung penilaian, pendapat dan hasil evaluasi. Isi poin-poin tadi akan dijadikan parameter untuk proses *labelling* dengan 3 annotator. Dengan total 458 data yang di split menjadi 2 yaitu data *testing* dan data *training* menggunakan ratio 20:80. Klasifikasi dilakukan sebanyak 2 kali, pertama proses klasifikasi SVM dilakukan tanpa algoritma SVD dan hanya menggunakan TF-IDF pada fitur ekstraksinya. Hasilnya evaluasi yang didapat nilai akurasi 83%, nilai presisi 83%, *recall* 69%, spesifisitas 91%, nilai *F-measure* 75%. Proses Kedua, fitur ekstraksi menggunakan TF-IDF dan SVD dihasilkan akurasi 80%, nilai presisi 73%, *recall* 81%, spesifisitas 80%, nilai *F-measure* 77%.

**Kata kunci :** Defamasi, *Support vector machine*, *Singular value decomposition*, TF-IDF, *Truncated svd*, *Confusion matrix*.

## ABSTRACT

CANDRA FEBRIAN TRI PAMBUDI. Detection of Defamation Elements on Twitter Social Media Using Support Vector Machine Classification. Skripsi. Faculty of Mathematics and Natural Sciences, State University of Jakarta. August 2023. Under the guidance of Mrs Dr Ria Arafiyah, M.Si. and Mr Drs Mulyono, M.Kom.

Defamation is an act that intentionally makes the good name or reputation of the person to be polluted or bad. This leads to character assassination which is a violation of human rights. This skripsi aims to build a defamation detection model with SVM classification on Twitter and SafeNet data, calculate SVM performance and determine the effect of TF-IDF and SVD on feature extraction on model performance. Based on SKB 3 Minister point g, actions that are said to be defamation are actions that contain elements of accusation. Meanwhile, defamation does not include narratives that contain judgements, opinions and evaluation results. The contents of these points will be used as parameters for the labelling process with 3 annotators. With a total of 458 data split into 2, namely testing data and training data using a ratio of 20:80. Classification was carried out twice, first SVM classification process was carried out without the SVD algorithm and only used TF-IDF in feature extraction. The evaluation results obtained accuracy value 83%, precision value 83%, recall 69%, specificity 91%, F-measure value 75%. Second process, feature extraction using TF-IDF and SVD resulted in 80% accuracy, 73% precision value, 81% recall, 80% specificity, 77% F-measure value.

**Keywords :** Defamation, Support vector machine, Singular value decomposition, TF-IDF, Truncated svd, Confusion matrix.

## DAFTAR ISI

LEMBAR PERSETUJUAN .....	i
LEMBAR PERNYATAAN .....	ii
HALAMAN PERSEMBAHAN.....	iii
KATA PENGANTAR.....	iv
ABSTRAK .....	vi
ABSTRACT .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN .....	xvi
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
<b>A. Latar Belakang .....</b>	<b>1</b>
<b>B. Rumusan Masalah.....</b>	<b>5</b>
<b>C. Batasan Masalah .....</b>	<b>6</b>
<b>D. Tujuan Penelitian .....</b>	<b>6</b>
<b>E. Manfaat Penelitian .....</b>	<b>6</b>
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>8</b>
<b>A. Pencemaran Nama Baik atau Defamasi.....</b>	<b>8</b>
<b>B. Undang - Undang Informasi dan Transaksi Elektronik .....</b>	<b>8</b>
<b>C. Surat Keputusan Bersama 3 (Tiga) Menteri .....</b>	<b>10</b>
<b>D. NLP (<i>Nature Language Processing</i>) .....</b>	<b>13</b>
<b>E. <i>Web Crawling Data</i> .....</b>	<b>13</b>
<b>F. <i>Librari Tweepy</i>.....</b>	<b>14</b>
<b>G. <i>Fleiss Kappa</i> .....</b>	<b>14</b>
<b>H. <i>Text Pre-Processing</i> .....</b>	<b>16</b>
1. <i>Parsing</i> .....	16
2. <i>Data Cleaning</i> .....	17
3. <i>Case Folding</i> .....	17
4. <i>Tokenizing/Lexical Analysis</i> .....	17
5. <i>Normalization</i> .....	17
6. <i>Filtering/Stopword Removal</i> .....	18

7. <i>Stemming</i> .....	18
<b>I. <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i></b> .....	<b>19</b>
<b>J. <i>Singular Value Decomposition</i></b> .....	<b>21</b>
<b>K. <i>Reduction Matrix</i></b> .....	<b>23</b>
<b>L. <i>Support Vector Machine (SVM)</i></b> .....	<b>24</b>
<b>M. <i>Bahasa Pemrograman Python</i></b> .....	<b>27</b>
<b>N. <i>Evaluasi</i></b> .....	<b>28</b>
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>32</b>
<b>A. <i>Tahapan Penelitian</i></b> .....	<b>32</b>
<b>B. <i>Identifikasi Masalah</i></b> .....	<b>33</b>
<b>C. <i>Studi Literatur</i></b> .....	<b>33</b>
<b>D. <i>Proses Pengumpulan Dataset</i></b> .....	<b>33</b>
1. <i>Objek Penelitian</i> .....	35
2. <i>Jenis Data</i> .....	35
3. <i>Sumber Data</i> .....	35
<b>E. <i>Parsing dan Merging</i></b> .....	<b>35</b>
<b>F. <i>Pelabelan Data</i></b> .....	<b>37</b>
<b>G. <i>Pre-Processing</i></b> .....	<b>38</b>
1. <i>Data Cleaning</i> .....	39
2. <i>Case Folding</i> .....	40
3. <i>Tokenizing</i> .....	40
4. <i>Normalization</i> .....	41
5. <i>Stopwords Removal</i> .....	42
6. <i>Stemming</i> .....	43
<b>H. <i>Ekstraksi Fitur Teks</i></b> .....	<b>43</b>
1. <i>Pembobotan TF-IDF</i> .....	44
2. <i>Dekomposisi SVD</i> .....	46
3. <i>Reduksi Truncated SVD</i> .....	46
4. <i>Dikalikan seluruh matriks</i> .....	46
<b>I. <i>Klasifikasi Support Vector Machine</i></b> .....	<b>46</b>
<b>J. <i>Analisis Hasil</i></b> .....	<b>47</b>
<b>K. <i>Dokumentasi</i></b> .....	<b>48</b>



L. Perangkat Penelitian.....	48
M. Jadwal Penelitian .....	49
<b>BAB IV .....</b>	<b>50</b>
<b>HASIL DAN PEMBAHASAN .....</b>	<b>50</b>
A. Pengumpulan Data.....	50
B. <i>Parsing dan Merging</i> .....	51
C. Pelabelan Data.....	54
D. <i>Pre-processing Text</i> .....	57
1. <i>Data Cleaning</i> .....	58
2. <i>Case Folding</i> .....	63
3. <i>Tokenizing</i> .....	64
4. <i>Normalization</i> .....	65
5. <i>Stopwords Removal</i> .....	66
6. <i>Stemming</i> .....	67
E. Pembobotan kata.....	69
F. Pendekomposisian Matrik <i>Singular</i> .....	76
G. Reduksi Matrik <i>Singular</i> .....	80
H. Klasifikasi.....	81
I. Hasil dan Evaluasi.....	85
<b>BAB V.....</b>	<b>91</b>
<b>KESIMPULAN DAN SARAN .....</b>	<b>91</b>
A. KESIMPULAN.....	91
B. SARAN.....	92
<b>DAFTAR PUSTAKA .....</b>	<b>93</b>
<b>LAMPIRAN.....</b>	<b>96</b>
<b>DAFTAR RIWAYAT HIDUP .....</b>	<b>179</b>