



SKRIPSI

**KOMPARASI KINERJA ALGORITMA SVM DAN
RF DALAM KLASIFIKASI SENTIMEN DENGAN
DETEKSI SARKASME PADA KOMENTAR
YOUTUBE**

MUHAMMAD HILMAN HABIB HABIBI
NPM 20082010004

DOSEN PEMBIMBING
Eka Dyar Wahyuni, S.Kom., M.Kom
Reisa Permatasari, S.T., M.Kom

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA
2025**

LEMBAR PENGESAHAN

**KOMPARASI KINERJA ALGORITMA SVM DAN RF DALAM
KLASIFIKASI SENTIMEN DENGAN DETEKSI SARKASME PADA
KOMENTAR YOUTUBE**

Oleh :


MUHAMMAD HILMAN HABIB HABIBI
NPM. 20082010004

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 23 Januari 2025.

Eka Dyar Wahyuni, S.Kom., M.Kom.
NIP. 19841201 2021212 005


..... (Pembimbing I)

Reisa Permatasari, S.T., M.Kom.
NIP. 19920514 2022032 007


..... (Pembimbing II)

Nur Cahyo Wibowo, S.Kom., M.Kom.
NIP. 19790317 2021211 002


..... (Ketua Penguji)

Agung Brastama Putra, S.Kom., M.Kom.
NIP. 19851124 2021211 003


..... (Anggota Penguji II)

Seftin Fitri Ana Wati, S.Kom., M.Kom.
NPT. 212199 10 320267


..... (Anggota Penguji III)

Mengetahui,
Dekan Fakultas Ilmu Komputer


.....

Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 1994032 001

LEMBAR PERSETUJUAN

**KOMPARASI KINERJA ALGORITMA SVM DAN RF DALAM
KLASIFIKASI SENTIMEN DENGAN DETEKSI SARKASME PADA
KOMENTAR YOUTUBE**

Oleh:

MUHAMMAD HILMAN HABIB HABIBI

NPM. 20082010004

Telah disetujui untuk mengikuti ujian skripsi

Menyetujui,

Dosen Pembimbing I



Eka Dyar Wahyuni, S.Kom., M.Kom.

NIP. 19841201 2021212 005

Dosen Pembimbing II



Reisa Permatasari, S.T., M.Kom.

NIP. 19920514 2022032 007

**Koordinator Program Studi Sistem Informasi
Fakultas Ilmu Komputer**



Agung Brastama Putra, S.Kom., M.Kom.

NIP. 19851124 2021211 003

SURAT PERNYATAAN ORISINALITAS

Yang bertandatangan di bawah ini:

Nama Mahasiswa : MUHAMMAD HILMAN HABIB HABIBI
Program Studi : Sistem Informasi
Dosen Pembimbing : Eka Dyar Wahyuni, S.Kom., M.Kom
Reisa Permatasari, S.T., M.Kom

dengan ini menyatakan bahwa isi sebagian maupun keseluruhan disertasi dengan judul:

KOMPARASI KINERJA ALGORITMA SVM DAN RF DALAM KLASIFIKASI SENTIMEN DENGAN DETEKSI SARKASME PADA KOMENTAR YOUTUBE

adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diizinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri. Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya, 3 Februari 2025
Yang Membuat Pernyataan,



Muhammad Hilman Habib Habibi

ABSTRAK

Nama Mahasiswa / NPM : Muhammad Hilman Habib Habibi / 20082010004
Judul Skripsi : Komparasi Kinerja Algoritma SVM Dan RF dalam Klasifikasi Sentimen dengan Deteksi Sarkasme pada Komentar Youtube
Dosen Pembimbing : 1. Eka Dyarwahyuni, S.Kom., M.Kom
2. Reisa Permatasari, S.T., M.Kom

YouTube adalah platform berbagi video yang sering diakses di Indonesia, terutama dalam diskusi terkait PILKADA 2024. Isu-isu yang muncul menjelang PILKADA memicu pro dan kontra, mendorong masyarakat untuk beropini di media sosial. Klasifikasi sentimen bertujuan mengelompokkan opini menjadi positif atau negatif, tetapi sering terhambat oleh sarkasme, yaitu ironi yang menyampaikan makna berlawanan dengan pernyataan eksplisit. Skripsi ini memperoleh data melalui *crawling* komentar YouTube, yang kemudian diproses dengan *cleaning*, *case folding*, dan *stemming*. Klasifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) dengan berbagai skenario, termasuk metode *holdout* (80:20 dan 70:30), teknik *resampling* (*Random Oversampling*/ROS dan *Random Undersampling*/RUS), serta pembobotan kata TF-IDF dan TF-ABS. Berdasarkan evaluasi, algoritma SVM dengan ROS dan TF-IDF pada rasio 80:20 memberikan hasil terbaik untuk klasifikasi sentimen dengan akurasi 0.80, sedangkan SVM dengan TF-IDF tanpa *resampling* pada rasio 80:20 mencapai 0.73 untuk klasifikasi sarkasme. Sementara itu, RF dengan ROS dan TF-ABS pada rasio 70:30 menghasilkan akurasi 0.73 untuk klasifikasi sentimen, dan RF dengan ROS dan TF-IDF pada rasio yang sama mencapai 0.69 untuk klasifikasi sarkasme. Hasil ini menunjukkan keandalan SVM dalam menangkap pola data yang kompleks, terutama dalam klasifikasi sentimen dengan deteksi sarkasme

Kata kunci: PILKADA, Sentimen, Sarkasme, *Support Vector Machine* (SVM), *Random Forest* (RF)

ABSTRACT

Nama Mahasiswa / NPM : Muhammad Hilman Habib Habibi / 20082010004
Judul Skripsi : Comparison of the Performance of SVM and RF Algorithms in Sentiment Classification with Sarcasm Detection on YouTube Comments
Dosen Pembimbing : 1. Eka Dyarwahyuni, S.Kom., M.Kom
2. Reisa Permatasari, S.T., M.Kom

YouTube is a video sharing platform that is often accessed in Indonesia, especially in discussions related to the 2024 PILKADA. Issues that arise ahead of the PILKADA trigger pros and cons, encouraging people to express their opinions on social media. Sentiment classification aims to group opinions into positive or negative, but is often hampered by sarcasm, which is irony that conveys the opposite meaning to the explicit statement. This thesis obtains data through crawling YouTube comments, which are then processed with cleaning, case folding, and stemming. Classification is carried out using the Support Vector Machine (SVM) and Random Forest (RF) algorithms with various scenarios, including the holdout method (80:20 and 70:30), resampling techniques (Random Oversampling/ROS and Random Undersampling/RUS), and TF-IDF and TF-ABS word weighting. Based on the evaluation, the SVM algorithm with ROS and TF-IDF at a ratio of 80:20 gave the best results for sentiment classification with an accuracy of 0.80, while SVM with TF-IDF without resampling at a ratio of 80:20 achieved 0.73 for sarcasm classification. Meanwhile, RF with ROS and TF-ABS at a ratio of 70:30 produced an accuracy of 0.73 for sentiment classification, and RF with ROS and TF-IDF at the same ratio achieved 0.69 for sarcasm classification. These results demonstrate the reliability of SVM in capturing complex data patterns, especially in sentiment classification with sarcasm detection.

Keywords: PILKADA, Sentiment, Sarcasm, Support Vector Machine (SVM), Random Forest (RF)

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Komparasi Kinerja Algoritma SVM Dan RF dalam Klasifikasi Sentimen dengan Deteksi Sarkasme pada Komentar Youtube”** dapat terselesaikan dengan baik.

Perjalanan penyusunan skripsi ini menjadi pengalaman yang penuh pembelajaran, tantangan, dan momen berharga yang tidak mungkin terlewat tanpa dukungan berbagai pihak. Oleh karena itu, penulis dengan tulus menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Orang tua penulis yaitu Almh. Ibu Tris Tyaningrum, Bapak Yuliono dan segenap keluarga penulis, terima kasih sebesar-besarnya telah senantiasa memberikan doa, dukungan, materi dan hiburan selama perjuangan menyelesaikan skripsi ini.
2. Ibu Eka Dyar Wahyuni, S.Kom., M.Kom. sebagai dosen pembimbing pertama yang telah memberikan waktu, dukungan, arahan dan memberikan support penulis dalam proses pengerjaan skripsi ini.
3. Ibu Reisa Permatasari, S.T., M.Kom. selaku dosen pembimbing kedua yang telah membimbing, memberikan saran dan mengarahkan penulis dalam proses pengerjaan skripsi ini.
4. Bapak Prisa Marga Kusumantara, S.Kom., M.Cs. dan Bapak Agung Brastama Putra, S.Kom., M.Kom. selaku Dosen Wali yang telah membantu administrasi penulis selama perkuliahan.
5. Bapak Agung Brastama Putra, S.Kom., M.Kom. selaku koordinator program studi Sistem Informasi UPN “Veteran” Jawa Timur yang selalu memberikan kemudahan bagi mahasiswanya.
6. Seluruh Bapak dan Ibu dosen Sistem Informasi yang telah memberikan banyak ilmu pengetahuan yang bermanfaat selama perkuliahan.
7. Ocha Taniya Brigidta yang selalu meluangkan waktu serta tenaganya dalam proses penyusunan skripsi dan telah memberikan support serta memberi semangat ketika penulis merasa down.
8. Teman-teman seperjuangan kuliah yang sering bersama di “Teras 28”, tempat yang menjadi saksi perjuangan dalam menyelesaikan tugas kuliah hingga skripsi.

9. Teman-teman seperjuangan Sistem Informasi 2020 yang telah memberikan dukungan, motivasi dan doa, sehingga skripsi ini dapat selesai tepat waktu.

Penulis menyadari bahwa di dalam penyusunan skripsi ini jauh dari kata sempurna. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini.

Akhirnya, dengan segala keterbatasan yang penulis miliki, semoga skripsi ini dapat memberikan manfaat nyata bagi perkembangan ilmu pengetahuan, khususnya di bidang sistem informasi

Surabaya, Januari 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN ORISINALITAS	vii
ABSTRAK	ix
KATA PENGANTAR	xiii
DAFTAR ISI	xv
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxiii
DAFTAR NOTASI	xxv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	5
1.5. Sistematikan Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1. Penelitian Terdahulu	7
2.2. Landasan Teori.....	11
2.2.1. Youtube	11
2.2.2. Text Mining.....	12
2.2.3. Pilkada.....	13
2.2.4. <i>Python</i>	14
2.2.5. Sentimen.....	15
2.2.6. Sarkasme	16
2.2.7. <i>Preprocessing Text</i>	17
2.2.8. <i>Support Vector Machine</i>	17
2.2.9. Random Forest	18
2.2.10. <i>Counfusion Matrix</i>	20
2.2.11. TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>)	20
2.2.12. TF-ABS (<i>Term Frequency-Absolute Frequency</i>)	21
2.2.13. <i>Fleiss Kappa</i>	22
2.2.14. <i>Flask</i>	22
BAB III METODOLOGI PENELITIAN	25

3.1.	Studi Literatur	25
3.2.	Analisis Kebutuhan	25
3.3.	Pengumpulan Data	26
3.4.	Pelabelan Data.....	26
3.5.	Eksplorasi Data	27
3.6.	Pembagian Data	27
3.7.	Preprocessing Data.....	27
3.8.	Perancangan Model.....	28
3.8.1	Pembagian Dataset Sentimen.....	29
3.8.2	Pembobotan Data dengan <i>Term Weighting</i>	29
3.8.3	Klasifikasi Sentimen	30
3.8.4	Pembagian Dataset Sarkasme	30
3.8.5	Klasifikasi Sarkasme	30
3.8.6	Evaluasi Model.....	30
3.8.7	Simpan Model Terbaik.....	31
3.9.	Penerapan Model.....	31
3.10.	Validasi Model	31
BAB IV HASIL DAN PEMBAHASAN		33
4.1.	Implementasi Kebutuhan	33
4.1.1.	Kebutuhan Data.....	33
4.1.2.	Kebutuhan Hardware dan Software	33
4.2.	Pengumpulan Data Komentar	33
4.3.	Pelabelan Data.....	35
4.3.1.	Pelabelan Sentimen Vader	38
4.3.2.	Pelabelan Sentimen TextBlob	39
4.3.3.	Pelabelan Sentimen Cardiffnlp	40
4.3.4.	Modus Hasil Pelabelan Sentimen.....	42
4.3.5.	Pelabelan Sarkasme GPT-3.5 Turbo	43
4.3.6.	Mengeluarkan Hasil Sentimen yang Berlabel Netral.....	45
4.4.	Eksplorasi Data Komentar	45
4.5.	Pembagian Data	48
4.6.	Preprocessing Data.....	48
4.6.1.	<i>Case Folding</i>	48
4.6.2.	<i>Stemming</i>	49

4.7.	Perancangan Model.....	49
4.7.1.	Pembagian Dataset Sentimen.....	50
4.7.2.	Pembobotan Data dengan <i>Term Weighting</i>	52
4.7.3.	Klasifikasi Sentimen	54
4.7.4.	Pembagian Dataset Sarkasme	62
4.7.5.	Klasifikasi Sarkasme	65
4.7.6.	Evaluasi Model.....	71
4.7.7.	Simpan Model Terbaik.....	78
4.8.	Penerapan Model.....	80
4.9.	Validasi Model	81
4.10.	Pembahasan.....	82
BAB V PENUTUP		85
5.1.	Kesimpulan	85
5.2.	Saran Pengembangan	86
DAFTAR PUSTAKA		87

DAFTAR GAMBAR

Gambar 1.1 Populasi dan Pengguna Media Sosial di Dunia.....	2
Gambar 2.1 Komentar Youtube	12
Gambar 2.2 Cara Kerja SVM.....	18
Gambar 2.3 Cara Kerja Random Forest	19
Gambar 3.1 Alur Metodologi Penelitian.....	25
Gambar 3.2 Alur Proses Pelabelan Data	26
Gambar 3.3 Tahap <i>Preprocessing</i> Data	28
Gambar 3.4 Perancangan Model	29
Gambar 4.1 <i>Source Code Crawling</i> Data.....	34
Gambar 4.2 Potongan <i>Source Code</i> Proses <i>Cleaning</i>	36
Gambar 4.3 <i>Source Code Translate</i> Data	37
Gambar 4.4 Potongan <i>Source Code</i> Pelabelan Vader	38
Gambar 4.5 Potongan <i>Source Code</i> Pelabelan TextBlob.....	39
Gambar 4.6 Potongan <i>Source Code</i> Pelabelan Cardiffnlp.....	41
Gambar 4.7 <i>Source Code</i> Modus untuk Sentimen <i>Result</i>	42
Gambar 4.8 Potongan <i>Source Code</i> Pelabelan Sarcasm GPT35.....	44
Gambar 4.9 <i>Source Code Drop</i> Label Netral.....	45
Gambar 4.10 <i>Source Code</i> EDA Profile Report	45
Gambar 4.11 Hasil <i>Shortcut</i> EDA <i>Profile Report</i>	46
Gambar 4.12 <i>Source Code Fleiss Kappa</i>	47
Gambar 4.13 Hasil Fleiss Kappa.....	47
Gambar 4.14 <i>Source Code</i> Pembagian Data	48
Gambar 4.15 <i>Source Code</i> Proses <i>Case Folding</i>	48
Gambar 4.16 <i>Source Code</i> Proses <i>Stemming</i>	49
Gambar 4.17 <i>Source Code Import</i> <i>Dataframe</i>	50
Gambar 4.18 <i>Source Code</i> Holdout Dataset Sentimen 80:20.....	50
Gambar 4.19 Potongan <i>Source Code Resampling</i> Data Sentimen.....	51
Gambar 4.20 <i>Source Code</i> Holdout Dataset Sentimen 70:30.....	51
Gambar 4.21 Potongan <i>Source Code Resampling</i> Data Sentimen.....	52
Gambar 4.22 <i>Source Code Term Weighting</i> TF-IDF	53
Gambar 4.23 <i>Source Code Term Weighting</i> TF-ABS.....	54
Gambar 4.24 <i>Source Code</i> Model SVM	55

Gambar 4.25 <i>Source Code</i> Model Random Forest	55
Gambar 4.26 Potongan <i>Source Code</i> Sentimen SVM IF-IDF.....	56
Gambar 4.27 Potongan <i>Source Code</i> Sentimen SVM TF-ABS.....	56
Gambar 4.28 Potongan <i>Source Code</i> Sentimen SVM ROS TF-IDF	57
Gambar 4.29 Potongan <i>Source Code</i> SVM ROS TF-ABS	57
Gambar 4.30 Potongan <i>Source Code</i> SVM RUS TF-IDF	58
Gambar 4.31 Potongan <i>Source Code</i> SVM RUS TF-ABS	58
Gambar 4.32 Potongan <i>Source Code</i> Sentiment RF TF-IDF.....	59
Gambar 4.33 Potongan <i>Source Code</i> RF TF-ABS	59
Gambar 4.34 Potongan <i>Source Code</i> RF TF-IDF dengan ROF.....	60
Gambar 4.35 Potongan <i>Source Code</i> RF TF-ABS dengan ROS	60
Gambar 4.36 Potongan <i>Source Code</i> Sentimen TF-IDF dengan RUS	61
Gambar 4.37 Potongan <i>Source Code</i> Sentimen TF-ABS dengan RUS.....	61
Gambar 4.38 <i>Source Code</i> Pembagian <i>Dataframe</i> Sarkasme.....	62
Gambar 4.39 <i>Source Code</i> Holdout Dataset Sarkasme 80:20	62
Gambar 4.40 Potongan <i>Source Code</i> <i>Resampling</i> Data Sarkasme 80:20	63
Gambar 4.41 <i>Source Code</i> Holdout Dataset Sarkasme 70:30	63
Gambar 4.42 Potongan <i>Source Code</i> <i>Resampling</i> Data Sarkasme 70:30	64
Gambar 4.43 Potongan <i>Source Code</i> Sarkasme SVM TF-IDF.....	65
Gambar 4.44 Potongan <i>Source Code</i> Sarkasme SVM TF-ABS	65
Gambar 4.45 Potongan <i>Source Code</i> Sarkasme SVM ROS TF-IDF.....	66
Gambar 4.46 Potongan <i>Source Code</i> Sarkasme SVM ROS TF-ABS	66
Gambar 4.47 Potongan <i>Source Code</i> Sarkasme SVM RUS TF-IDF.....	67
Gambar 4.48 Potongan <i>Source Code</i> Sarkasme SVM RUS TF-ABS	68
Gambar 4.49 Potongan <i>Source Code</i> Sarkasme RF TF-IDF	68
Gambar 4.50 Potongan <i>Source Code</i> Sarkasme RF TF-ABS	69
Gambar 4.51 Potongan <i>Source Code</i> Sarkasme RF ROS TF-IDF.....	69
Gambar 4.52 Potongan <i>Source Code</i> Sarkasme RF ROS TF-ABS	70
Gambar 4.53 Potongan <i>Source Code</i> Sarkasme RF RUS TF-IDF.....	70
Gambar 4.54 Potongan <i>Source Code</i> Sarkasme RF RUS TF-ABS	71
Gambar 4.55 Actual Label dan Predicted Label Sentimen	72
Gambar 4.56 Actual Label dan Predicted Label Sarkasme	72
Gambar 4.57 <i>Source Code</i> Penyimpanan Model	78
Gambar 4.58 <i>Source Code</i> memuat model terbaik untuk <i>pipeline</i>	79

Gambar 4. 59 <i>Source Code</i> membuat <i>pipeline</i> sentimen dan sarkasme	79
Gambar 4.60 <i>Source Code</i> Menggabungkan Kedua <i>Pipeline</i>	80
Gambar 4.61 Potongan <i>Source Code</i> Penerapan Model	81
Gambar 4.62 Hasil Pembuatan <i>Website</i>	81
Gambar 4. 63 Proses Validasi Model.....	82

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	7
Tabel 2.2 Confusion Matrix	20
Tabel 4.1 Hasil Crawling Data	35
Tabel 4.2 Hasil Proses Cleaning	36
Tabel 4.3 Hasil Translate	37
Tabel 4.4 Hasil Pelabelan Vader	39
Tabel 4.5 Hasil Pelabelan TextBlob	40
Tabel 4.6 Hasil Pelabelan Cardiff	41
Tabel 4.7 Hasil Sentiment Result	42
Tabel 4.8 Hasil Pelabelan Sarcasm GPT35	44
Tabel 4.9 Skala Penilaian <i>Fleiss Kappa</i>	46
Tabel 4.10 Hasil Skenario Evaluasi Model	73
Tabel 4.11 Evaluasi Model Skenario Terbaik	77

DAFTAR NOTASI

TP	:	<i>True Positive</i> , Jumlah data yang benar-benar positif
TN	:	<i>True Negative</i> , Jumlah data yang benar-benar negatif
FP	:	<i>False Positive</i> , Jumlah Data yang sebenarnya negatif
FN	:	<i>False Negative</i> , Jumlah data yang sebenarnya positif
$TF-IDF(t, d)$:	Merupakan bobot istilah (term) t dalam dokumen d
$TF(t, d)$:	Term Frequency (TF) mengukur seberapa sering istilah t muncul dalam dokumen d
$IDF(t)$:	Mengukur seberapa unik atau spesifik istilah t dalam koleksi dokumen (corpus).
N	:	Total jumlah dokumen dalam corpus.
n_t	:	Jumlah dokumen yang mengandung istilah t
df	:	Jumlah dokumen yang mengandung kata yang dicari.