

**PERBANDINGAN ALGORITMA SVM DAN RANDOM FOREST
UNTUK ANALISIS SENTIMEN CPNS DI MEDIA SOSIAL X**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HANAFI ASNAN

20.11.3481

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PERBANDINGAN ALGORITMA SVM DAN RANDOM FOREST
UNTUK ANALISIS SENTIMEN CPNS DI MEDIA SOSIAL X**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HANAFI ASNAN

20.11.3481

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN ALGORITMA SVM DAN RANDOM FOREST
UNTUK ANALISIS SENTIMEN CPNS DI MEDIA SOSIAL X**

yang disusun dan diajukan oleh

Hanafi Asnan

20.11.3481

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Februari 2025

Dosen Pembimbing,



Majid Rahardi, S.Kom., M.Eng

NIK. 190302393

HALAMAN PENGESAHAN
SKRIPSI
PERBANDINGAN ALGORITMA SVM DAN RANDOM FOREST
UNTUK ANALISIS SENTIMEN CPNS DI MEDIA SOSIAL X

yang disusun dan diajukan oleh

Hanafi Asnan

20.11.3481

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Anna Baita, S.Kom., M.Kom.
NIK. 190302290

Dr. Kumara Ari Yuana, ST , MT
NIK. 190302575

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom. Ph.D
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Hanafi Asnan
NIM : 20.11.3481

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN ALGORITMA SVM DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN CPNS DI MEDIA SOSIAL TWITTER

Dosen Pembimbing : Majid Rahardi, S.Kom., M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari 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 tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian 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.

Yogyakarta, 20 Februari 2025

Yang Menyatakan,



Hanafi Asnan

HALAMAN PERSEMBAHAN

Skripsi ini dipersembahkan kepada mereka yang telah berjasa dan menjadi sumber inspirasi dalam hidup saya. Rasa terima kasih yang mendalam disampaikan kepada semua yang telah membimbing, memberikan semangat, serta menanamkan ilmu yang akhirnya berkembang menjadi karya ilmiah ini. Dalam berbagai keadaan, mereka adalah pilar kekuatan yang tak tergantikan.

1. Segala puji bagi Allah SWT, atas rahmat dan petunjuk-Nya, penulis dapat menyelesaikan skripsi ini. Rasa syukur dan pengabdian sepenuhnya dipersembahkan kepada-Nya, sebagai sumber segala kebijaksanaan dan petunjuk.
2. Terima kasih yang tak terhingga kepada kedua orang tua, yang dengan penuh kasih sayang, doa, serta dukungan baik secara moral maupun material, menjadi kekuatan utama bagi penulis dalam menyelesaikan skripsi ini.
3. Penghargaan yang sebesar-besarnya disampaikan kepada Bapak Majid Rahardi, S.Kom., M.Eng ,atas kesabaran, bimbingan, serta wawasan ilmiahnya yang sangat berharga dalam memberikan arahan bagi penulis.
4. Ucapan terima kasih juga ditujukan kepada seluruh teman-teman Informatika 03, yang selalu memberikan semangat, berbagi pengalaman, serta menjadi rekan diskusi yang berarti. Kalian adalah dukungan sosial yang sangat berharga.

KATA PENGANTAR

Bagian Dengan segala rasa syukur kepada Allah, Yang Maha Pengasih dan Maha Penyayang, penulis menyampaikan puji dan syukur atas nikmat serta rahmat-Nya yang melimpah. Laporan skripsi ini berhasil diselesaikan sebagai syarat penyelesaian studi Strata Satu (S-1) di Program Studi Informatika Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Selama proses penulisan ini, penulis tidak hanya mendapatkan berbagai bantuan dan bimbingan, tetapi juga masukan berharga dari berbagai pihak yang turut mendukung. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan penghargaan dan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. Sebagai Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta M.Kom. Ph.D. sebagai Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya Duhita, M.Kom. sebagai Kaprodi S1 Informatika Universitas Amikom Yogyakarta.
4. Bapak Majid Rahardi, S.Kom., M.Eng sebagai Dosen Pembimbing.
5. Bapak dan Ibu Dosen di Universitas Amikom Yogyakarta yang telah memberikan bimbingan dan ilmu pengetahuan selama proses perkuliahan.
6. Kepada seluruh teman-teman dan pihak-pihak yang tidak bisa disebutkan satu persatu.

Yogyakarta, 20 February 2025

Penulis

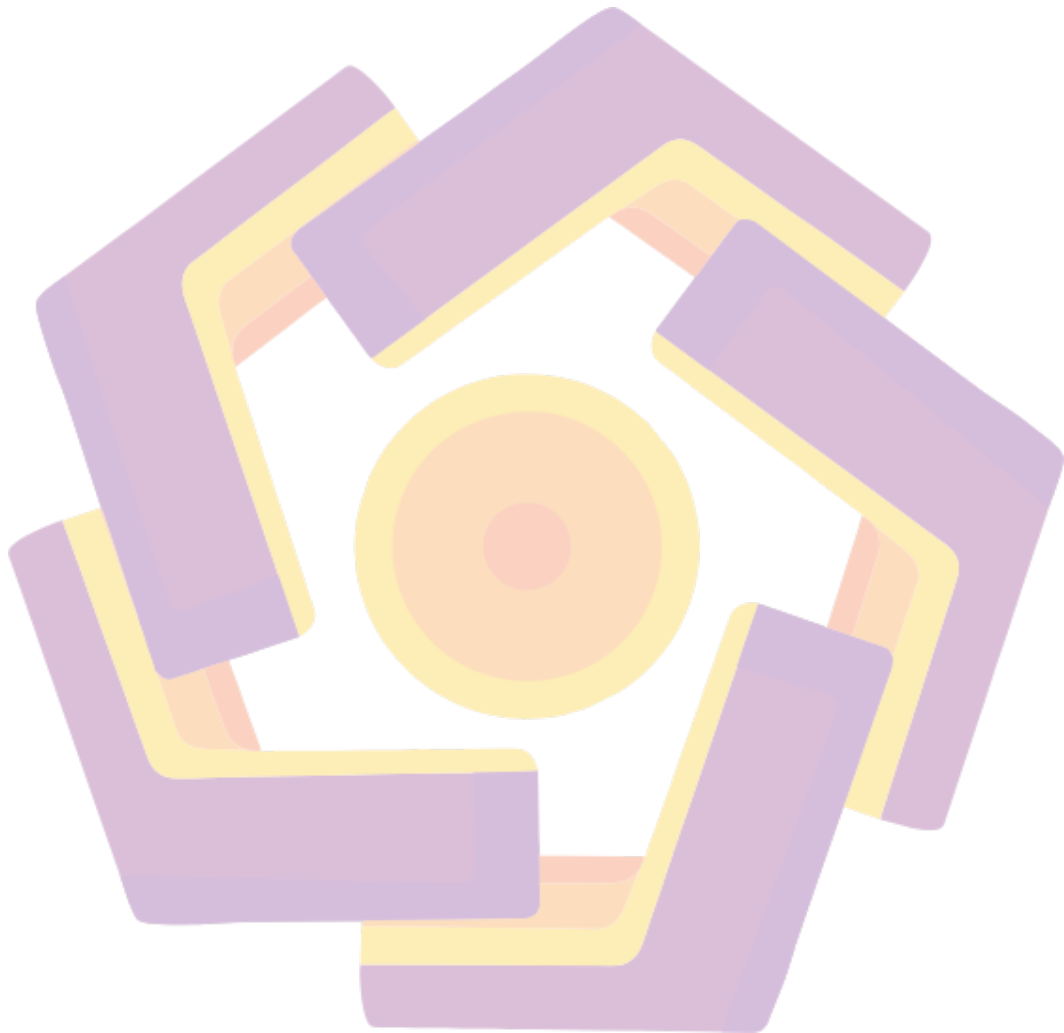
DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiv
INTISARI.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA.....	5

2.1 Studi Literatur	5
2.2 Dasar Teori	11
2.2.1 Analisis sentimen	11
2.2.2 Crawling Data	12
2.2.3 Pre-Processing.....	12
2.2.4 Supervised Learning	14
2.2.5 Python	14
2.2.6 Data Train-Test Split.....	15
2.2.7 SVM.....	15
2.2.8 Random Forest	16
2.2.9 Confusion Matrix	18
BAB III METODE PENELITIAN	20
3.1 Objek Penelitian.....	20
3.2 Alur Penelitian	20
3.2.1 Crawling Data	21
3.2.2 Import Dataset.....	22
3.2.3 Pre-procesing	22
3.2.4 Representasi fitur	29
3.2.5 Split Data.....	30
3.2.6 Model Training	30
3.2.7 Testing Data	31

3.2.8	Evaluasi.....	31
3.3	Alat dan Bahan.....	32
BAB IV HASIL DAN PEMBAHASAN		33
4.1	Proses Crawling data pada X	33
4.1.1	Instalasi Crawling Data.....	34
4.1.2	Konfigurasi pencarian pada Media sosial X	35
4.1.3	Hasil Crawling Data.....	36
4.1.4	Penggabungan data	37
4.2	Pembahasan sintaks program.....	39
4.2.1	Pembahasan sintaks python	39
4.3	Evaluasi Model	47
4.3.1	Random Forest	47
4.3.2	SVM.....	52
4.4	Hasil perbandingan	56
4.5	Penjelasan Hasil Perbandingan.....	57
4.5.1	Hasil Akurasi.....	57
4.5.2	Precision.....	57
4.5.3	Recall	58
4.5.4	F1-Score	59
BAB V PENUTUP.....		60
5.1	Kesimpulan	60

5.2 Saran	60
REFERENSI	61



DAFTAR TABEL

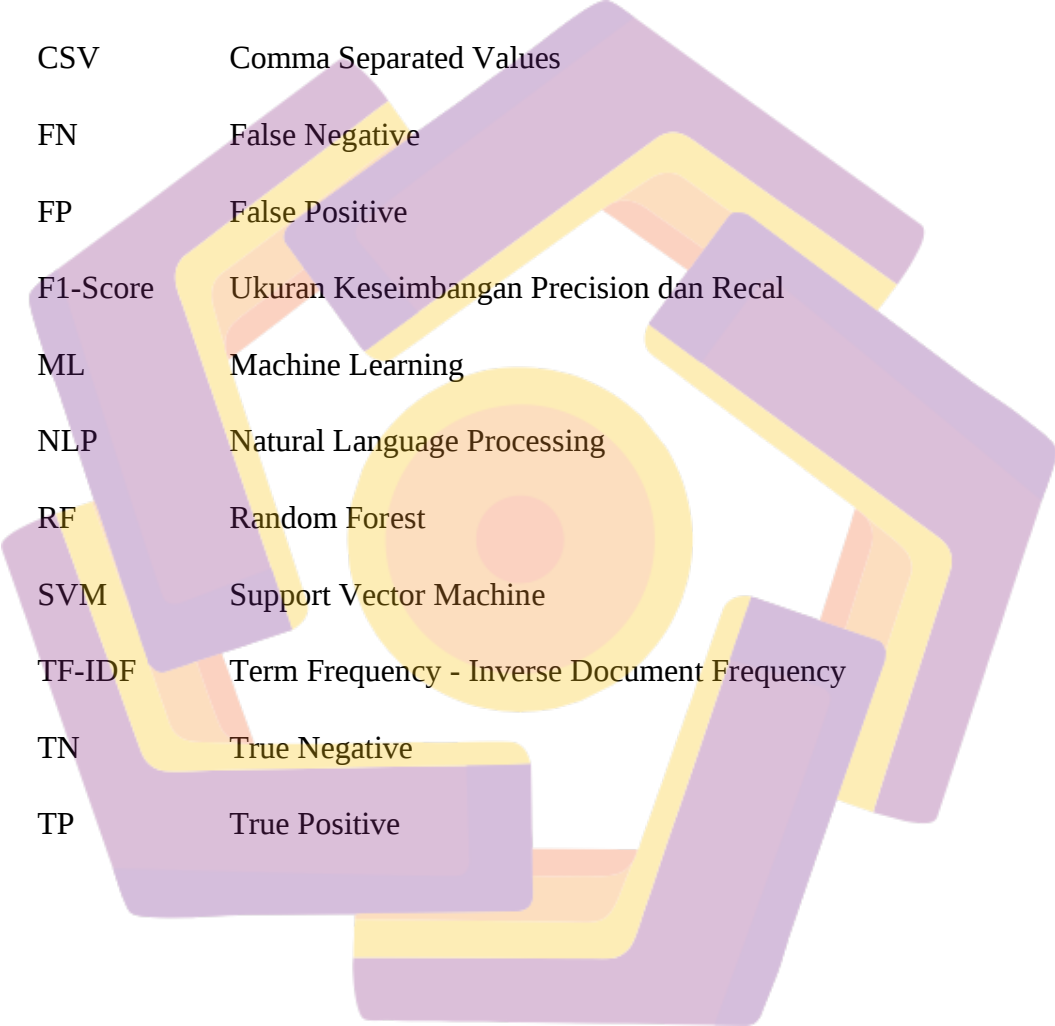
Tabel 2.1 Keaslian Penelitian	9
Tabel 2.2 Model Confusion Matrix	20
Tabel 3.1 Perbandingan Menghapus Spasi	26
Tabel 3.2 Perbandingan Menghapus Tanda Baca dan Angka	27
Tabel 3.3 Perbandingan Mengganti Huruf Kapital ke Huruf Normal	28
Tabel 3.4 Perbandingan Menghapus URL	29
Tabel 3.5 Alat dan Bahan	35
Tabel 3.6 Perangkat Lunak	36
Tabel 4.5 perhitungan Random Forest	48
Tabel 4.6 perhitungan SVM	59
Tabel 4.7 Hasil Perbandingan Precision	65
Tabel 4.8 Hasil Perbandingan Recall	66
Tabel 4.9 Hasil Perbandingan F1-Score	67

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	24
Gambar 3.2 Normalization	30
Gambar 3.3 Tokenization	30
Gambar 3.4 Stopwords	31
Gambar 3.5 Stemming	32
Gambar 3.6 Mengubah Sentimen ke Angka	32
Gambar 3.7 Jumlah Data Berdasarkan Sentimen	33
Gambar 4.1 Instalasi Node.js	38
Gambar 4.2 Instalasi Tweet Harvest	38
Gambar 4.3 Instalasi Pandas	39
Gambar 4.4 Auth Token	40
Gambar 4.5 Code Sebelum Melakukan Crawling Data	40
Gambar 4.6 Hasil Crawling Data	41
Gambar 4.7 Penggabungan Data	42
Gambar 4.8 Hasil Penggabungan Data	43
Gambar 4.9 Import Library	44
Gambar 4.10 Code untuk Menghapus Spasi Berlebih	45
Gambar 4.11 Code untuk Menghapus Tanda Baca dan Angka	46
Gambar 4.12 Code untuk Mengubah Huruf Besar Jadi Kecil	46
Gambar 4.13 Stopwords	47
Gambar 4.14 Normalization	48

Gambar 4.15 Tokenization	49
Gambar 4.16 stemming	49
Gambar 4.17 Code untuk Menghapus URL	50
Gambar 4.18 Code untuk mengubah sentimen ke angka	51
Gambar 4.19 Code untuk pemisahan data training dan testing	51
Gambar 4.20 Code TF-IDF	52
Gambar 4.21 Matrix Random Forest	53
Gambar 4.22 Matrix SVM	58
Gambar 4.23 Visualisasi Bar Random Forest	64
Gambar 4.24 Visualisasi Bar SVM	64

DAFTAR LAMBANG DAN SINGKATAN



AI	Artificial Intelligence (Kecerdasan Buatan)
API	Application Programming Interface
CPNS	Calon Pegawai Negeri Sipil
CSV	Comma Separated Values
FN	False Negative
FP	False Positive
F1-Score	Ukuran Keseimbangan Precision dan Recall
ML	Machine Learning
NLP	Natural Language Processing
RF	Random Forest
SVM	Support Vector Machine
TF-IDF	Term Frequency - Inverse Document Frequency
TN	True Negative
TP	True Positive

INTISARI

Sentimen Analisis pada media sosial semakin berkembang seiring meningkatnya interaksi pengguna terhadap berbagai isu publik, termasuk Calon Pegawai Negeri Sipil (CPNS). Penelitian ini bertujuan untuk membandingkan kinerja algoritma Support Vector Machine (SVM) dan Random Forest dalam menganalisa sentimen terkait CPNS di X. Data yang digunakan dikumpulkan melalui Crawling data dari X Menggunakan kata kunci #CPNS dalam rentang waktu dari Bulan Agustus sampai November. Setelah data diperoleh, dilakukan tahap pre-processing seperti pembersihan teks, tokenisasi, stopword removal, stemming, dan representasi fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Data kemudian dibagi menjadi data latih dan uji dengan perbandingan 80:20.

Evaluasi model dilakukan dengan menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model SVM memiliki akurasi sebesar 81%, sedikit lebih tinggi dibandingkan Random Forest yang memiliki akurasi 80%. SVM lebih unggul dalam mendeteksi sentimen negatif dan positif, sementara Random Forest memiliki performa lebih baik dalam mengklasifikasikan sentimen netral. Namun, kedua model masih mengalami kesulitan dalam mengenali sentimen netral dengan tingkat recall yang relatif rendah.

Penelitian ini menunjukkan bahwa pemilihan algoritma dalam analisis sentimen bergantung pada tujuan klasifikasi dan distribusi data sentimen yang digunakan. Hasil dari penelitian ini dapat menjadi referensi bagi pengembangan model analisis sentimen yang lebih akurat dan efektif dalam memahami opini publik terhadap kebijakan pemerintah.

Kata Kunci: Analisis Sentimen, CPNS, Support Vector Machine, Random Forest, X.

ABSTRACT

Abstract Sentiment analysis on social media is increasingly developing along with increasing user interaction on various public issues, including the selection of Candidates for Civil Servants (CPNS). This research aims to compare the performance of the Support Vector Machine (SVM) and Random Forest algorithms in analyzing sentiment related to CPNS on X. The data used was collected through crawling data from X using the keyword #CPNS in the time period from August to November. After the data is obtained, pre-processing stages are carried out such as text cleaning, tokenization, stopword removal, stemming, and feature representation using Term Frequency-Inverse Document Frequency (TF-IDF). The data is then divided into training and test data with a ratio of 80:20.

Model evaluation is carried out using accuracy, precision, recall and F1-score metrics. Experimental results show that the SVM model has an accuracy of 81%, slightly higher than Random Forest which has an accuracy of 80%. SVM is superior in detecting negative and positive sentiments, while Random Forest has better performance in classifying neutral sentiments. However, both models still experience difficulty in recognizing neutral sentiment with relatively low recall rates.

This research shows that the choice of algorithm in sentiment analysis depends on the classification objectives and distribution of the sentiment data used. The results of this research can be a reference for developing sentiment analysis models that are more accurate and effective in understanding public opinion towards government policies.

Keywords: Sentiment Analysis, CPNS, Support Vector Machine, Random Forest, X.