

**ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP
PENDAFTARAN CPNS MENGGUNAKAN NAIVE BAYES
CLASSIFIER DAN SUPPORT VECTOR MACHINE**

SKRIPSI

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



disusun oleh

MAUDI PRAMESTI DEWI

20.11.3505

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS SENTIMEN PENGGUNA YOTUBE TERHADAP
PENDAFTARAN CPNS MENGGUNAKAN NAIVE BAYES
CLASSIFIER DAN SUPPORT VECTOR MACHINE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

MAUDI PRAMESTI DEWI

20.11.3505

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP
PENDAFTARAN CPNS MENGGUNAKAN NAIVE BAYES CLASSIFIER
DAN SUPPORT VECTOR MACHINE**

yang disusun dan diajukan oleh

Maudi Pramesti Dewi

20.11.3505

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Desember 2025

Dosen Pembimbing,


Mulia Sulistiyono, S.Kom., M.Kom
NIK. 190302248

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP
PENDAFTARAN CPNS MENGGUNAKAN NAIVE BAYES DAN
SUPPORT VECTOR MACHINE**

yang disusun dan diajukan oleh

Maudi Pramesti Dewi

20.11.3505

Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Rifda Faticha Alfa Aziza, M.Kom
NIK. 190302392

Novi Prisma Yunita, M.Kom
NIK. 190302526

Mulia Sulistiyono, S.Kom., M.Kom
NIK. 190302248



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 12 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, S.Kom., M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Maudi Pramesti Dewi
NIM : 20.11.3505

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP
PENDAFTARAN CPNS MENGGUNAKAN NAIVE BAYES CLASSIFIER
DAN SUPPORT VECTOR MACHINE**

Dosen Pembimbing: Mulia Sulistiyono, S.Kom., M.Kom.

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, 12 Desember 2025

Yang Menyatakan,



Maudi Pramesti Dewi

HALAMAN PERSEMBAHAN

Dengan mengucap rasa Syukur yang mendalam atas kehadiran Allah SWT, yang telah memberikan hidayah – Nya atas kekuatan dan kesahatan penulis, sehingga skripsi ini dapat terselesaikan dengan baik dan tanpa terkendala. Oleh karena itu, penulis ingin mempersembahkan kepada:

1. Allah SWT atas segala limpahan rahmat, karunia, kesehatan, dan kemudahan yang diberikan selama proses menyusun dan menyelesaikan skripsi ini.
2. Bapak dan Ibu, Sutarno dan Sumini terima kasih atas doa, semangat, nasehat, dukungan, serta kasih sayang hingga saat ini.
3. Keluarga besar tersayang, yang selalu senantiasa memberikan semangat, dukungan dan juga doa restu dalam setiap langkah perjalanan ini.
4. Aditya Galih Prasetya yang telah memberikan doa, semangat serta motivasi yang tidak berhenti kepada penulis. Terimakasih sudah ikut serta dalam proses penulis untuk menyelesaikan skripsi ini, selalu percaya pada kemampuan saya dan sudah menjadi tempat saya berkeluh kesah selama ini.
5. Teman – teman kelas penulis yang saling berbagi pengalaman, pengetahuan dan memberikan dukungan sehingga skripsi ini dapat terselesaikan yang tidak dapat penulis sebutkan satu persatu.
6. Kepada diri sendiri, terimakasih sudah mampu bertahan dalam kondisi dan situasi apapun, sehingga dapat menyelesaikan skripsi ini dengan baik.

KATA PENGANTAR

Puji Syukur penulis panjatkan kepada Allah SWT atas nikmat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul **“ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP PENDAFTARAN CPNS MENGGUNAKAN NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE ”**. Skripsi ini sebagai salah satu syarat untuk memperoleh gelar sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, Proses menyelesaikan skripsi ini benar-benar tidak bisa lepas dari bantuan, dukungan, saran, dan bahkan kritik yang diberikan oleh banyak orang di sekitar saya. Karena itu, saya ingin secara khusus mendedikasikan hasil kerja ini kepada mereka semua, sebagai bentuk terima kasih atas segala kontribusi yang telah membuat perjalanan ini lebih mudah dan bermakna. Tanpa mereka, mungkin saja saya masih tersesat di tengah tumpukan buku dan data. Oleh karena itu penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusrini, M.Kom, selaku Dekan Fakultas Ilmu Komputer di Universitas AMIKOM Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom, selaku Ketua Program Studi Informatika di Universitas AMIKOM Yogyakarta.
4. Bapak Mulia Sulistiyono, M.Kom, selaku Dosen Pembimbing yang dengan sabar dan ikhlas meluangkan waktu, tenaga, dan pikiran, serta memberikan arahan, motivasi dan saran-saran yang sangat berharga kepada penulis selama proses penyusunan skripsi ini.
5. Bapak Mulia Sulistiyono, S.Kom., M.Kom. selaku Dewan Penguji atas segala masukan, kritik, dan saran yang telah diberikan sehingga membantu dalam penyempurnaan skripsi ini

6. Seluruh pihak yang telah mendukung penulisan skripsi ini yang tidak dapat penulis sebutkan satu persatu.

Penulis sepenuhnya menyadari bahwa masih terdapat banyak kekurangan dalam penulisan skripsi ini, baik dari aspek metodologi, analisis data, maupun kedalaman pembahasan. Oleh karena itu, saya sangat mengharapkan kritik dan saran yang konstruktif dari para pembaca, dosen, serta pakar terkait, guna melengkapi serta memperkaya informasi yang telah saya tuangkan di dalam karya ini, serta untuk memperbaiki dan mengembangkan penelitian selanjutnya. Semoga skripsi ini dapat memberikan kontribusi yang bermakna bagi berbagai pihak, termasuk mahasiswa, akademisi, dan praktisi di bidang studi ini, serta mendorong kemajuan ilmu pengetahuan. Terima kasih atas perhatian dan dukungannya.

Yogyakarta, 12 Desember 2025

Penulis

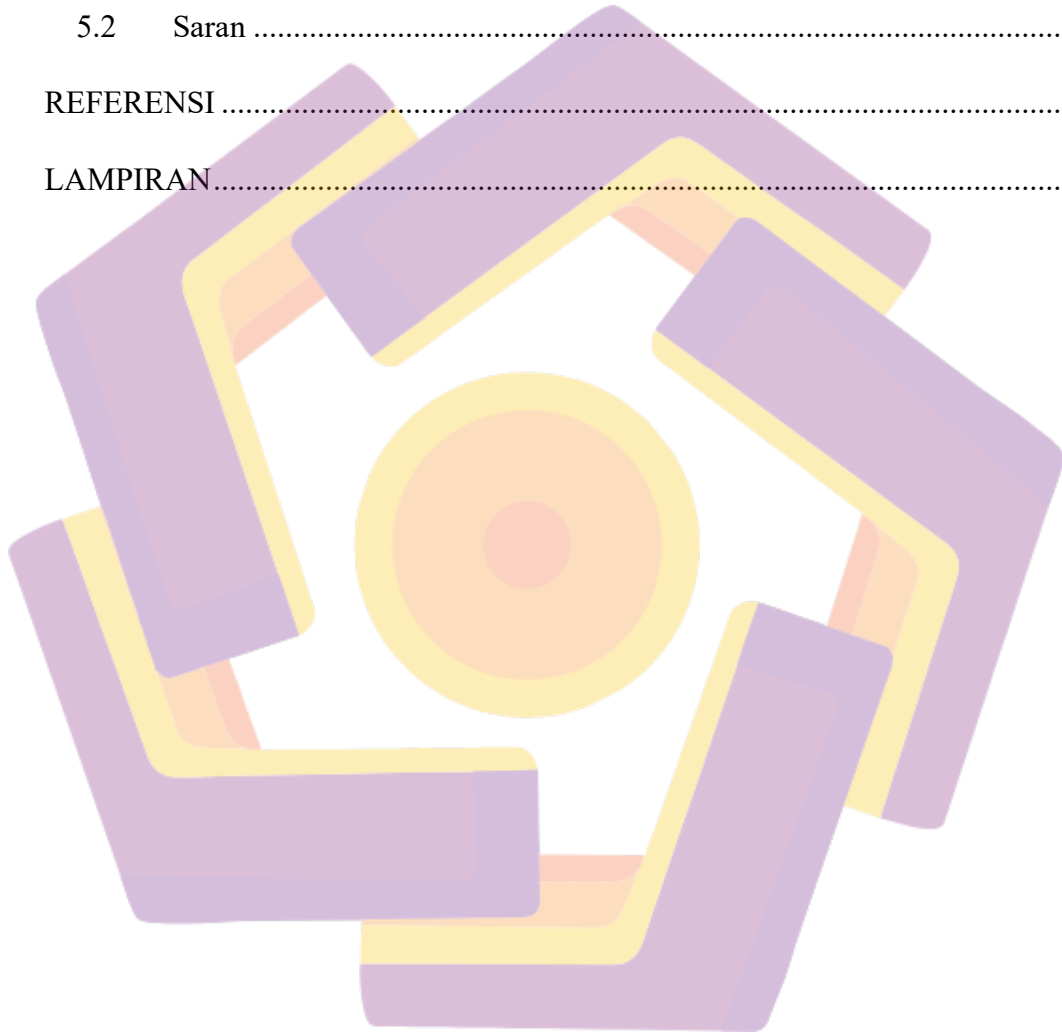
DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xvi
DAFTAR LAMBANG DAN SINGKATAN	xvii
DAFTAR ISTILAH	xviii
INTISARI	xix
<i>ABSTRACT</i>	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3

BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	14
2.2.1 Youtube.....	14
2.2.2 Analisis Sentimen	14
2.2.3 <i>Preprocessing Data</i>	14
2.2.4 <i>Text Mining</i>	16
2.2.5 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	16
2.2.6 <i>Algoritma Naïve Bayes</i>	17
2.2.7 <i>Algoritma Support Vector Machine</i>	18
2.2.8 <i>Kernel Radial Basis Function (RBF)</i>	18
2.2.9 <i>Python</i>	19
2.2.10 <i>Google Collaboratory</i>	19
2.2.11 <i>GridSearch CV</i>	19
2.2.12 <i>Confusion Matrix</i>	20
2.2.13 <i>Split Data 80:20</i>	21
BAB III METODE PENELITIAN	22
3.1 Objek Penelitian.....	22
3.2 Alur Penelitian	22
3.2.1 Pengumpulan Data	22
3.2.2 <i>Preprocessing Data</i>	23
3.2.3 Pelabelan Data.....	29

3.2.4	<i>Splitting Data</i>	29
3.2.5	<i>Balancing Data</i>	29
3.2.6	Modeling	29
3.2.7	Evaluasi Model	30
3.2.8	Kesimpulan	30
3.3	Alat dan Bahan	30
BAB IV HASIL DAN PEMBAHASAN		32
4.1	Pengumpulan Data	32
4.2	<i>Preprocessing</i>	32
4.2.1	<i>Cleanning</i>	32
4.2.2	<i>Case Folding</i>	35
4.2.3	<i>Normalization</i>	35
4.2.4	<i>Tokenization</i>	37
4.2.5	<i>Stopword Removal</i>	38
4.2.6	<i>Stemming</i>	38
4.3	Pelabelan Data	39
4.4	Frekuensi Kata	41
4.5	N-Gram Frekuensi Kata	44
4.6	Splitting Dataset	47
4.7	Term Frequency-Inverse Document Frequency (TF-IDF)	48
4.8	Algoritma Naïve Bayes Multinomial	50
4.9	Algoritma Support Vector Machine Kernel RBF	51

4.10 Synthetic Minority Over-sampling Technique (SMOTE)	53
4.11 Visualisasi	56
BAB V PENUTUP	61
5.1 Kesimpulan	61
5.2 Saran	62
REFERENSI	63
LAMPIRAN	64



DAFTAR TABEL

Tabel 3. 1 Hasil <i>Cleaning Data</i>	23
Tabel 3. 2 Hasil <i>Case Folding</i>	24
Tabel 3. 3 Kamus Kata Baku	25
Tabel 3. 4 Hasil Normalisasi.....	25
Tabel 3. 5 Hasil <i>Tokenizing</i>	26
Tabel 3. 6 Hasil <i>Stopword Removal</i>	27
Tabel 3. 7 Hasil <i>Stemming Data</i>	28
Tabel 3. 8 Tabel <i>Hardware</i>	30
Tabel 3. 9 Tabel <i>Software</i>	31
Tabel 4. 1 Skor TF-IDF tertinggi.....	49

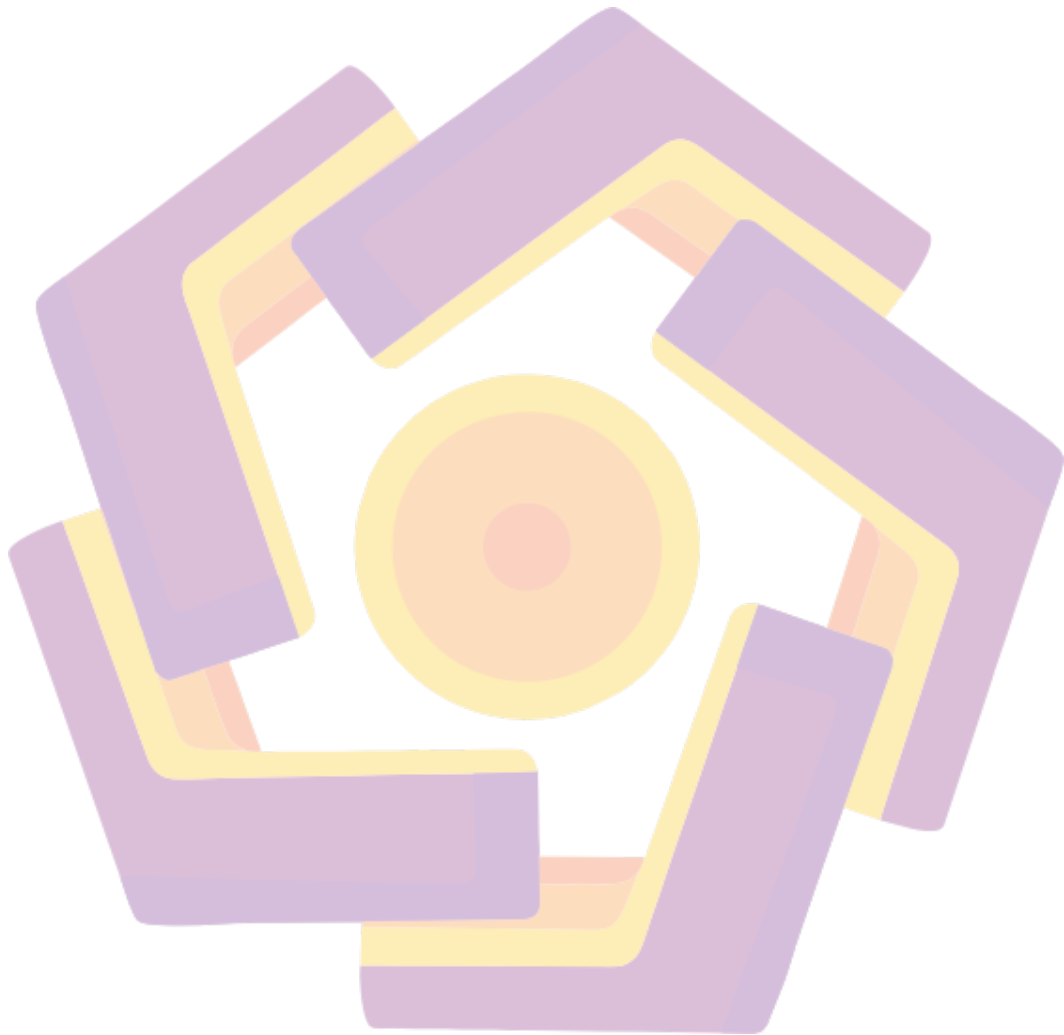
DAFTAR GAMBAR

Gambar 4. 1 Hasil <i>Scraping</i> dari Apify	32
Gambar 4. 2 <i>Source Code Cleanning</i>	34
Gambar 4. 3 Hasil <i>Cleaning</i>	34
Gambar 4. 4 <i>Source Code Case Folding</i>	35
Gambar 4. 5 Hasil <i>Case Folding</i>	35
Gambar 4. 6 <i>Source Code Normalization</i>	37
Gambar 4. 7 Hasil <i>Normalization</i>	37
Gambar 4. 8 <i>Source Code Tokenization</i>	37
Gambar 4. 9 Hasil <i>Tokenization</i>	38
Gambar 4. 10 <i>Source Code Stopword Removal</i>	38
Gambar 4. 11 Hasil <i>Stopword Removal</i>	38
Gambar 4. 12 <i>Source Code Stemming Data</i>	39
Gambar 4. 13 Hasil <i>Stemming Data</i>	39
Gambar 4. 14 <i>Source Code Pelabelan Data</i>	40
Gambar 4. 15 Hasil dari Pelabelan Data.....	40
Gambar 4. 16 Hasil Jumlah Analisis Sentimen Data	41
Gambar 4. 17 <i>Source Code Frekuensi Kata</i>	42
Gambar 4. 18 Hasil dari Sentimen Negatif dan Positif di Frekuensi Kata	42
Gambar 4. 19 <i>Source Code</i> Yang Menampilkan 20 Frekuensi Kata Teratas	43
Gambar 4. 20 Hasil dari Menampilkan 20 Frekuensi Kata Teratas.....	43

Gambar 4. 21 <i>Source Code</i> N-Gram Frekuensi Kata Untuk Sentimen Positif dan Negatif dari 10 Frekuensi Kata Teratas	45
Gambar 4. 22 Hasil Sentimen Positif dan Negatif di N-Gram Frekuensi Kata dari 10 Frekuensi Kata Teratas.....	45
Gambar 4. 23 <i>Source Code</i> N-Gram Frekuensi Kata Untuk Visualisasi Kata	46
Gambar 4. 24 Hasil Visualisasi Kata Untuk N-Gram Frekuensi Kata.....	46
Gambar 4. 25 <i>Source Code</i> <i>Splitting Dataset</i>	48
Gambar 4. 26 Hasil Jumlah Data Latih dan Data Uji Dataset	48
Gambar 4. 27 <i>Source Code</i> untuk <i>TF-IDF</i>	49
Gambar 4. 28 <i>Source Code</i> <i>Naive Bayes Multinomial</i>	50
Gambar 4. 29 Hasil Akurasi <i>Naïve Bayes Multinomial</i>	51
Gambar 4. 30 <i>Source Code</i> <i>Support Vector Machine</i>	52
Gambar 4. 31 Hasil Akurasi <i>Support Vector Machine</i>	52
Gambar 4. 32 <i>Source Code</i> Setelah <i>SMOTE</i>	55
Gambar 4. 33 Hasil Setelah di <i>SMOTE</i> dengan rasio 0.95	56
Gambar 4. 34 Hasil <i>Confusion Matrix SVM</i>	57
Gambar 4. 35 Hasil <i>Confusion Matrix Multinomial NB</i>	58
Gambar 4. 36 <i>WordCloud</i> Sentimen Negatif	59
Gambar 4. 37 <i>WordCloud</i> Sentimen Positif	59

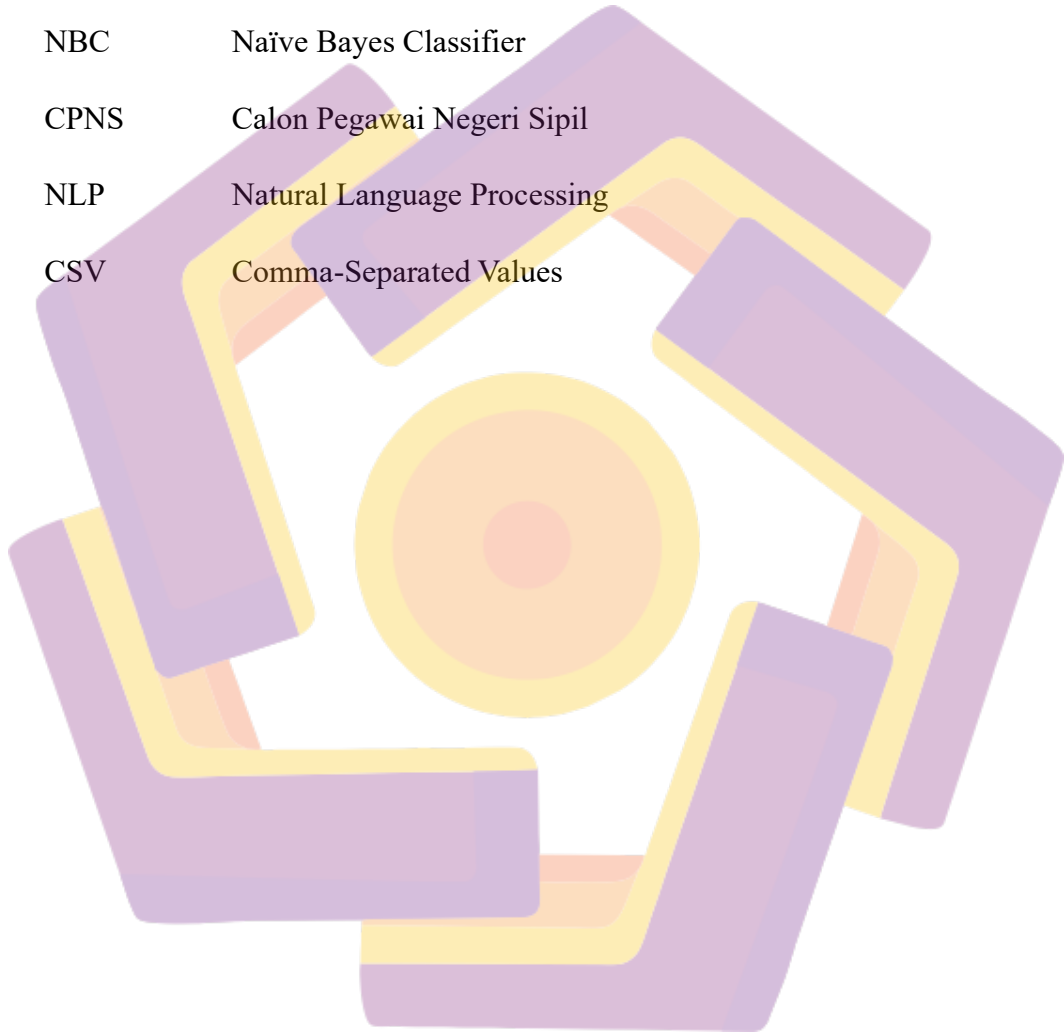
DAFTAR LAMPIRAN

Lampiran 1 Link Dataset	64
Lampiran 2 Link <i>Google Collab</i>	64



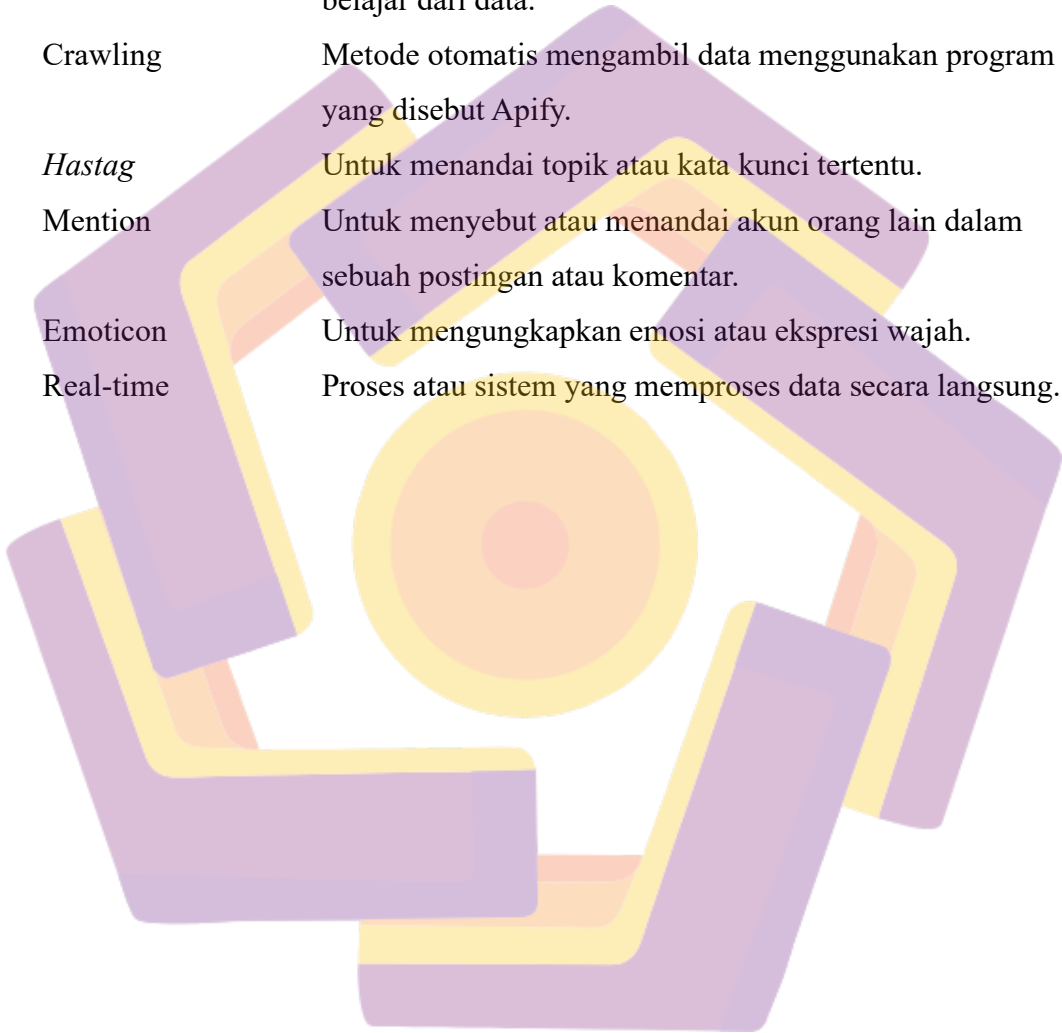
DAFTAR LAMBANG DAN SINGKATAN

NLTK	Natural Language Toolkit
URL	Uniform Resource Locator
SVM	Support Vector Machine
NBC	Naïve Bayes Classifier
CPNS	Calon Pegawai Negeri Sipil
NLP	Natural Language Processing
CSV	Comma-Separated Values



DAFTAR ISTILAH

Dataset	Kumpulan data yang terorganisir yang digunakan sebagai dasar dalam analisis.
Machine Learning	Cabang dari kecerdasan buatan yang memungkinkan untuk belajar dari data.
Crawling	Metode otomatis mengambil data menggunakan program yang disebut Apify.
<i>Hashtag</i>	Untuk menandai topik atau kata kunci tertentu.
Mention	Untuk menyebut atau menandai akun orang lain dalam sebuah postingan atau komentar.
Emoticon	Untuk mengungkapkan emosi atau ekspresi wajah.
Real-time	Proses atau sistem yang memproses data secara langsung.



INTISARI

Pendaftaran Calon Pegawai Negeri Sipil (CPNS) menjadi isu publik yang selalu mendapat perhatian luas, termasuk melalui komentar pengguna pada platform YouTube. Tingginya volume opini positif dan negatif sering kali membuat persepsi publik sulit dipetakan secara objektif, sehingga berpotensi menimbulkan interpretasi keliru mengenai transparansi proses seleksi dan kualitas penyelenggaraan. Ketidakmampuan pihak terkait dalam membaca kecenderungan opini publik dapat memengaruhi kepercayaan masyarakat terhadap sistem rekrutmen pemerintah. Oleh karena itu, dibutuhkan analisis sentimen yang sistematis untuk mengetahui respons nyata masyarakat terhadap proses pendaftaran CPNS.

Penelitian ini menggunakan pendekatan text mining dengan tahapan pengumpulan komentar, pembersihan teks, normalisasi, tokenisasi, dan pembobotan menggunakan TF-IDF. Dua algoritma pembelajaran mesin diuji, yaitu Support Vector Machine (SVM) dengan kernel RBF dan Multinomial Naive Bayes (MNB). Hyperparameter tuning diterapkan untuk meningkatkan performa model. Data dibagi menjadi data latih dan data uji, kemudian dievaluasi menggunakan metrik akurasi, precision, recall, dan confusion matrix untuk memvalidasi kinerja model secara empiris.

Hasil evaluasi menunjukkan bahwa SVM RBF menghasilkan performa terbaik dengan akurasi 85,16%, melampaui MNB yang memperoleh akurasi 74,48%. Temuan ini menegaskan bahwa SVM lebih efektif dalam memetakan pola non-linear pada teks komentar YouTube terkait pendaftaran CPNS. Penelitian ini bermanfaat bagi pemerintah, lembaga seleksi, dan peneliti kebijakan publik untuk memahami persepsi masyarakat secara lebih akurat. Ke depan, pengembangan dapat dilakukan dengan menambah sumber data lintas platform serta mengintegrasikan model deep learning guna meningkatkan ketepatan analisis.

Kata kunci: Analisis Sentimen; YouTube; Pendaftaran CPNS 2024; Naive Bayes Classifier; Support Vector Machine; Pembelajaran Mesin; Text Mining; Analisis Opini Publik; Data Media Sosial.

ABSTRACT

The recruitment of Civil Servant Candidates (CPNS) has consistently attracted widespread public attention, particularly through user comments on YouTube. The high volume of positive and negative opinions often makes it difficult to objectively map public perception, potentially leading to misinterpretations regarding the transparency and quality of the selection process. The inability of stakeholders to accurately capture the dynamics of public sentiment may undermine public trust in the government's recruitment system. Therefore, a systematic sentiment analysis is required to identify the real responses expressed by the public toward the CPNS registration process.

This study applies a text mining approach involving the stages of comment collection, text cleaning, normalization, tokenization, and weighting using TF-IDF. Two machine learning algorithms were evaluated, namely Support Vector Machine (SVM) with an RBF kernel and Multinomial Naive Bayes (MNB), with hyperparameter tuning applied to enhance model performance. The dataset was divided into training and testing sets, and model effectiveness was assessed using accuracy, precision, recall, and confusion matrix metrics.

The evaluation results show that the SVM RBF model achieved the best performance with an accuracy of 85.16%, outperforming MNB, which achieved 74.48%. These findings indicate that SVM is more effective in capturing non-linear patterns within YouTube comment texts related to CPNS registration. This research is beneficial for government institutions, selection organizers, and public policy researchers in understanding public perceptions more accurately. Future studies may incorporate multi-platform data sources and integrate deep learning methods to enhance analytical precision.

Keyword : YouTube Sentiment Analysis; Civil Servant Registration (CPNS) 2024; Naive Bayes Classifier; Support Vector Machine; Machine Learning; Text Mining; Public Opinion Analysis; Social Media Analytics.