

**PERBANDINGAN ANALISIS SENTIMEN MENGGUNAKAN
COMPLEMENT NAÏVE BAYES DAN SUPPORT VECTOR
MACHINE**

SKRIPSI

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



disusun oleh
TIYO MARGI
20.11.3538

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024

**PERBANDINGAN ANALISIS SENTIMEN MENGGUNAKAN
COMPLEMENT NAÏVE BAYES DAN SUPPORT VECTOR
MACHINE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
TIYO MARGI
20.11.3538

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN ANALISIS SENTIMEN MENGGUNAKAN
COMPLEMENT NAÏVE BAYES DAN SUPPORT VECTOR MACHINE**

yang disusun dan diajukan oleh

Tiyo Margi

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Agustus 2024

Dosen Pembimbing,



Ike Verawati, M.Kom
NIK. 190302237

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN ANALISIS SENTIMEN MENGGUNAKAN
COMPLEMENT NAÏVE BAYES DAN SUPPORT VECTOR
MACHINE

yang disusun dan diajukan oleh

Tiyo Margi

20.11.3538

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Agustus 2024

Susunan Dewan Penguji

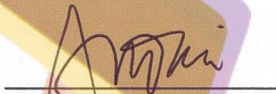
Nama Penguji

Ria Andriani, M.Kom
NIK. 190302458

Melany Mustika Dewi, M.Kom
NIK. 190302455

Ike Verawati, M.Kom
NIK. 190302237

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Agustus 2024

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 : Tiyo Margi
NIM : 20.11.3538

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN ANALISIS SENTIMEN MENGGUNAKAN COMPLEMENT NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

Dosen Pembimbing : Ike Verawati, 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, 20 Agustus 2024

Yang Menyatakan,



Tiyo Margi

HALAMAN PERSEMBAHAN

Dengan segala puji Syukur kepada Tuhan yang Maha Esa dan atas dukungan dan doa dari orang-orang tercinta, akhirnya skripsi ini dapat dirampungkan dengan baik dan tepat pada waktunya. Oleh karena itu, dengan rasa bangga dan bahagia saya ucapkan rasa Syukur dan terima kasih saya kepada:

1. Allah SWT, atas segala rahmat, hidayah dan karunia-Nya yang tiada terhingga, yang telah memberikan kekuatan dan kemampuan sehingga skripsi ini dapat di selesaikan
2. Kedua orang tua tercinta, Bapak Gimin A. Noor, SE dan Ibu Marsini yang selalu memberikan cinta, doa, dukungan, dan kasih sayang tanpa batas, serta menjadi sumber kekuatan dalam setiap langkah yang saya ambil.
3. Saudara dan keluarga besar, yang selalu memberikan semangat, dorongan, dan kehangatan yang tak tergantikan.
4. Para dosen dan pembimbing, yang telah memberikan ilmu, bimbingan, serta motivasi dalam menyelesaikan skripsi ini. Terima kasih atas segala arahan dan dukungannya.
5. Teman-teman seperjuangan, khususnya Agi Rizki Ananda dan Dandi Wicaksana yang selalu berbagi pengalaman, tawa, dan semangat selama masa perkuliahan hingga menyelesaikan skripsi ini.
6. Almamater tercinta, yang telah menjadi tempat saya menimba ilmu dan mengembangkan diri.

Semoga skripsi ini dapat memberikan manfaat dan menjadi referensi yang berguna bagi pembaca dan pengembangan ilmu pengetahuan.

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "PERBANDINGAN ANALISIS SENTIMEN MENGGUNAKAN COMPLEMENT NAÏVE BAYES DAN SUPPORT VECTOR MACHINE" ini dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Dalam penyusunan skripsi ini, penulis menyadari bahwa tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak, skripsi ini tidak akan terselesaikan dengan baik. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

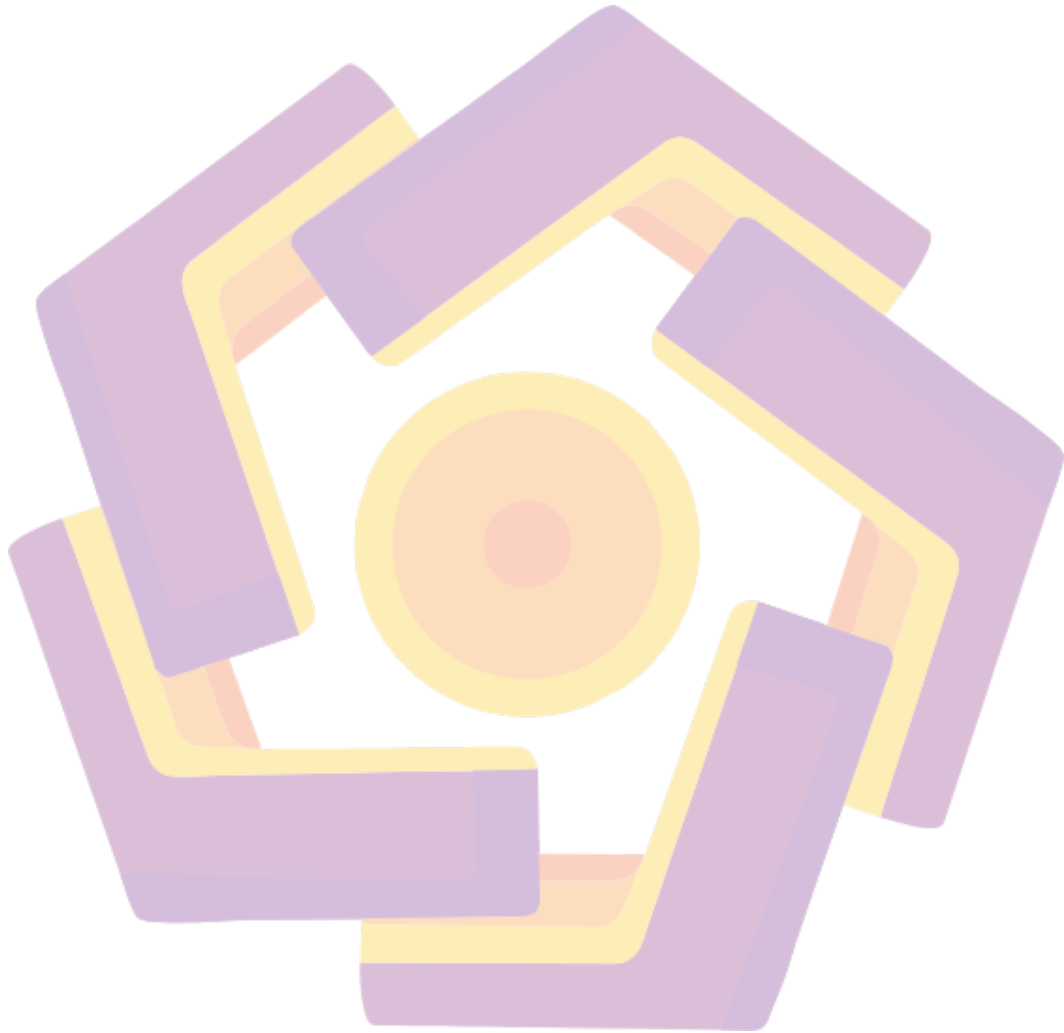
1. Ibu Ike Verawati, M.Kom., selaku dosen pembimbing, yang dengan penuh kesabaran, memberikan bimbingan, arahan, dan motivasi selama proses penyusunan skripsi ini.
2. Dosen Penguji, yang telah memberikan kritik, saran, serta masukan yang membangun demi kesempurnaan skripsi ini.
3. Kedua orang tua tercinta, yang selalu memberikan doa, dukungan moral dan material, serta semangat yang tiada henti.
4. Teman-teman seperjuangan, yang senantiasa berbagi pengetahuan, pengalaman, dan semangat selama masa studi hingga terselesaikannya skripsi ini.
5. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah membantu penulis, baik secara langsung maupun tidak langsung, dalam penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis dengan lapang hati menerima kritik dan saran yang membangun. Semoga skripsi ini dapat memberikan manfaat dan menjadi referensi yang berguna bagi pembaca dan pengembangan ilmu pengetahuan.

Akhir kata, penulis berharap semoga Allah SWT senantiasa memberikan rahmat dan ridho-Nya kepada kita semua.

Yogyakarta, 20 Agustus 2024

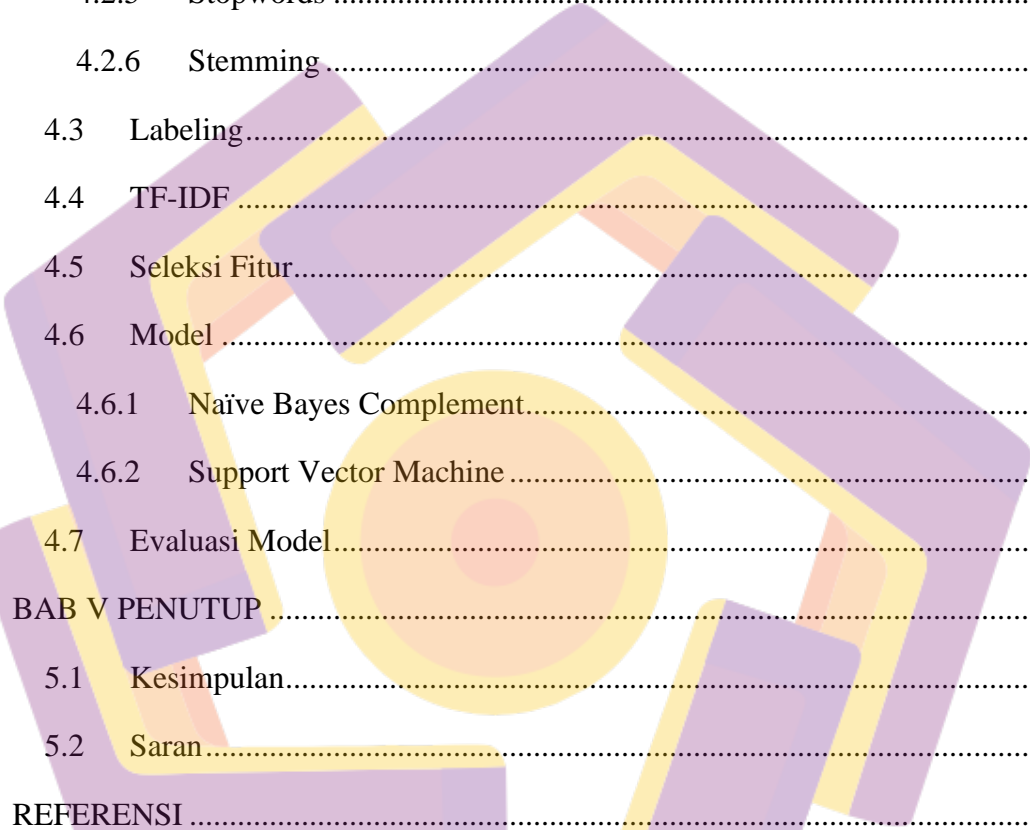
Penulis



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5

2.2	Dasar Teori	12
2.2.1	Text mining	12
2.2.2	Analisis Sentimen	12
2.2.3	Youtube	12
2.2.4	Text Preprocessing	13
2.2.5	TF-IDF	13
2.2.6	Chi-Square	13
2.2.7	Information Gain	14
2.2.8	Naïve Bayes Classifier	15
2.2.9	Support Vector Machine	16
2.2.10	Evaluasi Model	17
BAB III METODE PENELITIAN		19
3.1	Objek Penelitian	19
3.2	Alur Penelitian	19
3.2.1	Pengumpulan Data	20
3.2.2	Preprocessing	20
3.2.3	Labeling	21
3.2.4	TF-IDF	21
3.2.5	Selection Feature	22
3.2.6	Modelling	22
3.2.7	Evaluasi Model	23
3.3	Alat dan Bahan	23
BAB IV HASIL DAN PEMBAHASAN		25
4.1	Pengumpulan Data	25
4.2	Preprocessing Data	26



4.2.1	Case Folding	27
4.2.2	Cleaning	28
4.2.3	Normalization	30
4.2.4	Tokenizing	31
4.2.5	Stopwords	32
4.2.6	Stemming	34
4.3	Labeling	36
4.4	TF-IDF	39
4.5	Seleksi Fitur	39
4.6	Model	40
4.6.1	Naïve Bayes Complement	40
4.6.2	Support Vector Machine	41
4.7	Evaluasi Model	43
BAB V PENUTUP		51
5.1	Kesimpulan	51
5.2	Saran	51
REFERENSI		52

DAFTAR TABEL

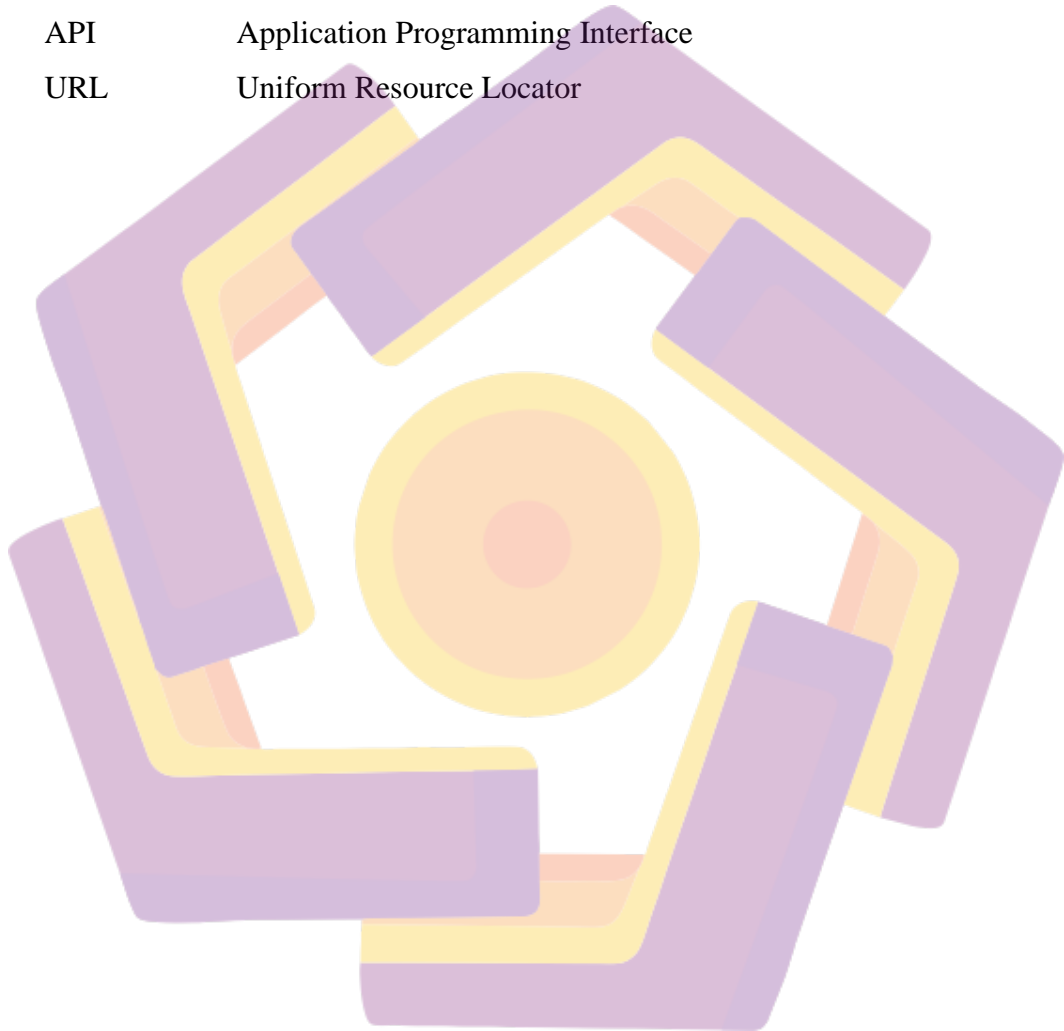
Tabel 2. 1 Keaslian Penelitian	8
Tabel 2. 2 Confusion Matrix	18
Tabel 3. 1 Data set	24
Tabel 4.1 Hasil Case Folding.....	27
Tabel 4. 2 Hasil Cleaning.....	29
Tabel 4. 3 Hasil Normalization.....	30
Tabel 4. 4 Hasil Tokenizing.....	32
Tabel 4. 5 Hasil Stopwords.....	33
Tabel 4. 6 Hasil Stemming.....	34
Tabel 4. 7 Hasil Preprocessing.....	35
Tabel 4.8 Hasil Labeling.....	37
Tabel 4.9 Frekuensi kemunculan kata.....	39
Tabel 4. 10 Hasil Chi-Square.....	39
Tabel 4.11 Hasil IG Score.....	40
Tabel 4. 12 Pembagian Rasio Data	40
Tabel 4. 13 Hasil Model CNB	41
Tabel 4. 14 Hasil Model SVM.....	42
Tabel 4. 15 Confusion Matrix CNB Rasio 70:30	43
Tabel 4. 16 Confusion Matrix SVM Rasio 70:30	43
Tabel 4. 17 Hasil Perbandingan CNB dan SVM pada Rasio 70:30.....	44
Tabel 4. 18 Confusion Matrix CNB Rasio 80:20	46
Tabel 4. 19 Confusion Matrix SVM Rasio 80:20	46
Tabel 4. 20 Hasil Perbandingan CNB dan SVM pada Rasio 80:20.....	47
Tabel 4. 21 Confusion Matrix CNB Rasio 90:10	48
Tabel 4. 22 Confusion Matrix SVM Rasio 90:10	49
Tabel 4. 23 Hasil Perbandingan CNB dan SVM pada Rasio 90:10.....	49

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	19
Gambar 4. 1 Source Code Crawling	26
Gambar 4. 2 Hasil Data Set.....	26
Gambar 4. 3 Source Code Case Folding.....	27
Gambar 4. 4 Source Code Cleaning.....	29
Gambar 4. 5 Source Code Normalization.....	30
Gambar 4. 6 Source Code Tokenizing.....	31
Gambar 4. 7 Source Code Stopwords.....	33
Gambar 4. 8 Source Code Stemming.....	34
Gambar 4. 9 Source Code Labeling.....	36
Gambar 4. 10 Hasil Sentimen.....	38
Gambar 4. 11 Grafik Perbandingan Model pada Rasio 70:30.....	45
Gambar 4. 12 Grafik Perbandingan Model pada Rasio 80:20.....	47
Gambar 4. 13 Grafik Perbandingan Model pada Rasio 90:10.....	50

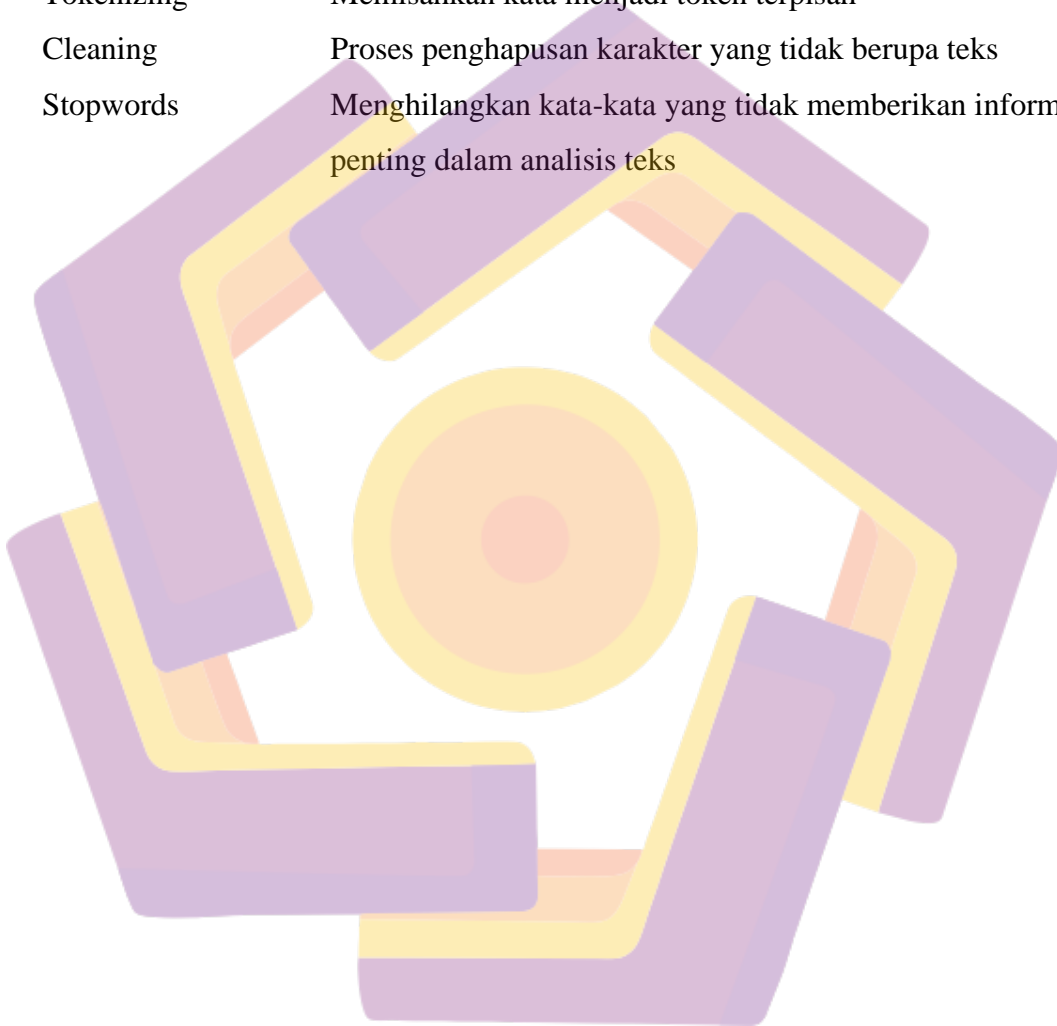
DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machine
CNB	Complement Naïve Bayes
TF-IDF	Term Frequency-Inverse Document Frequency
VADER	Valence Aware Dictionary and Sentiment Reasoner
API	Application Programming Interface
URL	Uniform Resource Locator



DAFTAR ISTILAH

Case Folding	Proses Mengubah semua data teks menjadi huruf kecil (lowercase)
Normalization	Tahap mengubah kata-kata tidak baku menjadi baku
Tokenizing	Memisahkan kata menjadi token terpisah
Cleaning	Proses penghapusan karakter yang tidak berupa teks
Stopwords	Menghilangkan kata-kata yang tidak memberikan informasi penting dalam analisis teks



INTISARI

Penelitian ini bertujuan untuk membandingkan kinerja dua metode klasifikasi yaitu Support Vector Machine (SVM) dan Naïve Bayes serta pengaruh seleksi fitur Chi-Square dan Information Gain dalam analisis sentiment komentar Youtube terhadap film Dirty Vote. Data dikumpulkan dari komentar Youtube sebanyak 2500 data. Hasil penelitian menunjukkan bahwa seleksi fitur meningkatkan recall dan F1-Score untuk model CNB, dengan akurasi stabil di kisaran 73-76%. Sebaliknya, untuk SVM, seleksi fitur memberikan sedikit variasi pada recall dan F1-Score, namun presisi tetap tinggi (88-89%). Akurasi SVM sedikit menurun pada rasio 70:30 dan 80:20, namun tetap stabil pada rasio 90:10. Secara keseluruhan, seleksi fitur lebih berdampak pada CNB dibandingkan SVM. SVM secara konsisten lebih efektif daripada CNB dalam mengklasifikasikan sentimen komentar, unggul dalam hal presisi, recall, F1-Score, dan akurasi di semua rasio data (70:30, 80:20, dan 90:10). Dengan akurasi mencapai 83% pada rasio 90:10 dan presisi tinggi (88-89%), SVM terbukti lebih unggul dibandingkan CNB, meskipun CNB menunjukkan beberapa peningkatan dengan penerapan seleksi fitur. Penelitian ini memberikan wawasan berharga mengenai pengaruh seleksi fitur dalam analisis sentimen, yang dapat bermanfaat bagi peneliti dalam memilih algoritma dan seleksi fitur yang tepat dan sebagai referensi untuk penelitian selanjutnya.

Kata kunci: Analisis Sentimen, Support Vector Machine, Complement Naïve Bayes, Information Gain, Chi-Square

ABSTRACT

This study aims to compare the performance of two classification methods, Support Vector Machine (SVM) and Naïve Bayes, and the impact of feature selection using Chi-Square and Information Gain in sentiment analysis of YouTube comments on the film Dirty Vote. The data consists of 2500 YouTube comments. The results show that feature selection improves recall and F1-Score for the CNB model, with accuracy remaining stable in the range of 73-76%. In contrast, for SVM, feature selection results in slight variations in recall and F1-Score, while precision remains high (88-89%). SVM accuracy slightly decreases at ratios of 70:30 and 80:20 but remains stable at a ratio of 90:10. Overall, feature selection has a greater impact on CNB compared to SVM. SVM consistently outperforms CNB in classifying sentiment comments, excelling in precision, recall, F1-Score, and accuracy across all data ratios (70:30, 80:20, and 90:10). With accuracy reaching 83% at a 90:10 ratio and high precision (88-89%), SVM proves to be superior to CNB, although CNB shows some improvement with feature selection. This study provides valuable insights into the impact of feature selection in sentiment analysis, which can be beneficial for researchers in choosing appropriate algorithms and feature selection methods, serving as a reference for future research.

Keyword: *Sentiment Analysis, Support Vector Machine, Complement Naïve Bayes, Information Gain, Chi-Square*