

**PERBANDINGAN METODE KLASIFIKASI SUPPORT
VECTOR MACHINE (SVM) DAN NAÏVE BAYES DALAM
ANALISIS SENTIMEN KOMENTAR YOUTUBE CNN
TENTANG KERETA CEPAT JAKARTA-BANDUNG (WHOOS)**

SKRIPSI

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



disusun oleh

DANDI WICAKSANA

20.11.3551

Kepada

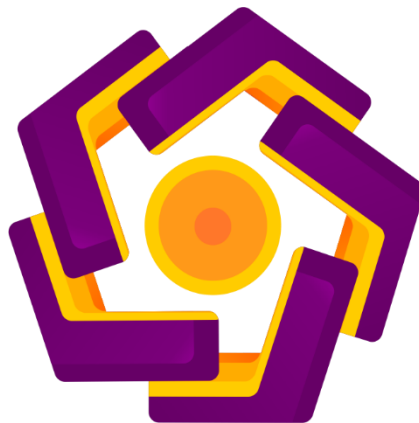
**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**COMPARISON OF SUPPORT VECTOR MACHINE (SVM)
AND NAÏVE BAYES CLASSIFICATION METHODS IN
SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON
CNN ABOUT THE JAKARTA-BANDUNG HIGH-SPEED
TRAIN (WHOOS)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

DANDI WICAKSANA

20.11.3551

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN METODE KLASIFIKASI SUPPORT VECTOR
MACHINE (SVM) DAN NAÏVE BAYES DALAM ANALISIS SENTIMEN
KOMENTAR YOUTUBE CNN TENTANG KERETA CEPAT JAKARTA-
BANDUNG (WHOOS)**


yang disusun dan diajukan oleh

Dandi Wicaksana

20.11.3551

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 10 Juli 2024

Dosen Pembimbing,


Nuri Cahyono, M.Kom
NIK. 190302278

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) DAN NAÏVE BAYES DALAM ANALISIS SENTIMEN KOMENTAR YOUTUBE CNN TENTANG KERETA CEPAT JAKARTA-BANDUNG (WHOOS)

yang disusun dan diajukan oleh

Dandi Wicaksana

20.11.3551

Telah dipertahankan di depan Dewan Penguji
pada tanggal 30 Juli 2024

Susunan Dewan Penguji

Nama Penguji

Lilis Dwi Farida, S.Kom, M.Eng
NIK. 190302288

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

Nuri Cahyono, M.Kom
NIK. 190302278

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 30 Juli 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 : Dandi Wicaksana
NIM : 20.11.3551

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan metode klasifikasi SUPPORT VECTOR MACHINE (SVM) dan NAÏVE BAYES dalam analisis sentimen komentar youtube CNN tentang kereta cepat Jakarta-Bandung (WHOOS)

Dosen Pembimbing : Nuri Cahyono 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, 30 Juli 2024

Yang Menyatakan,



Dandi Wicaksana

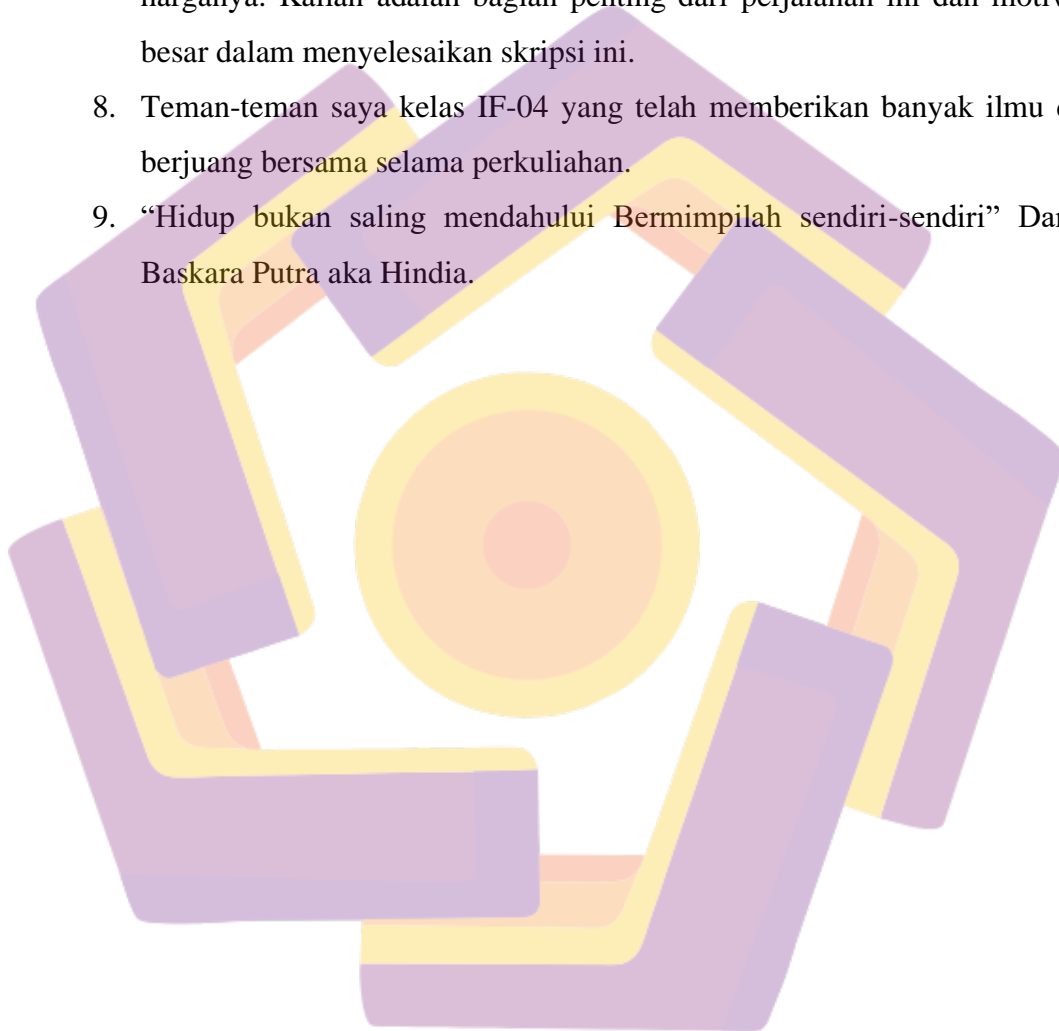
HALAMAN PERSEMBAHAN

Dengan rasa syukur Alhamdulillahirobbil'alamiin, segala puji bagi Allah Subhanahu Wa Ta'ala yang telah mencurahkan Rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Perbandingan metode klasifikasi SUPPORT VECTOR MACHINE (SVM) dan NAÏVE BAYES dalam analisis sentimen komentar youtube CNN tentang kereta cepat Jakarta-Bandung (WHOOS)”

Penulis mempersembahkan skripsi ini kepada :

1. Seseorang yang sangat berarti bagi penulis, Bapak Suharjo dan Ibu Suparyani yang selalu memberikan cinta, dukungan, dan doa yang tiada henti. Terima kasih atas segala pengorbanan, motivasi, dan kasih sayang yang kalian berikan selama ini. Kalian adalah sumber kekuatan terbesar dalam hidupku.
2. **fa inna ma'al-'usri yusr, inna ma'al-'usri yusrā**
Maka, sesungguhnya beserta kesulitan ada kemudahan
(Al-Insyirah · Ayat 5-6)
3. Kedua Kakak ku beserta pasangannya, Danang Santoso dan Dian Putri Nastiti beserta Deni Wiyanti dan Febri Yudi Wibowo yang selalu selalu menjadi pilar kekuatan dan sumber inspirasi. Terima kasih atas dukungan, semangat, dan kasih sayang yang tiada henti. Kehadiran kalian sangat berarti dalam setiap pencapaian hidupku.
4. Danurendra Bagaskara Putra Santoso dan Muhammad Takiyan Wibowo Yang kehadirannya membawa keceriaan dan kebahagiaan dalam hidupku. Semoga kelak kalian dapat mencapai cita-cita yang tinggi dan membanggakan keluarga. Kehadiran kalian selalu menjadi penyemangat tersendiri dalam menyelesaikan skripsi ini.
5. Keluarga penulis yang tidak bisa disebutkan satu persatu, terimakasih sudah membantu doa dan memberikan semangat sehingga penulis bisa menyelesaikan skripsi ini.

6. Kepada Bapak Nuri Cahyono yang telah memberikan bimbingan dan ilmu baru dalam pengerjaan skripsi.
7. Kepada sahabat saya Wangsit Bintang Darmawan, Agi Riski Ananda, Tiyo Margi dan Raniya Rakarahayu yang selalu ada di setiap suka dan duka. Terima kasih atas dukungan, canda tawa, dan kebersamaan yang tak ternilai harganya. Kalian adalah bagian penting dari perjalanan ini dan motivasi besar dalam menyelesaikan skripsi ini.
8. Teman-teman saya kelas IF-04 yang telah memberikan banyak ilmu dan berjuang bersama selama perkuliahan.
9. “Hidup bukan saling mendahului Bermimpilah sendiri-sendiri” Daniel Baskara Putra aka Hindia.



KATA PENGANTAR

Dengan memanjatkan puja dan puji Syukur kehadiran Tuhan Yang Maha Esa, atas rahmat yang melimpah serta kesehatan, sehingga penulis bisa menyelesaikan skripsi ini dengan judul “Perbandingan metode klasifikasi SUPPORT VECTOR MACHINE (SVM) dan NAÏVE BAYES dalam analisis sentimen komentar youtube CNN tentang kereta cepat Jakarta-Bandung (WHOOS)”, sebagai salah satu syarat untuk menyelesaikan program strata 1 Sarjana Ilmu Komputer di Universitas AMIKOM Yogyakarta.

Selama proses menyusun skripsi ini, penulis mendapat dukungan, bantuan, bimbingan, dan nasehat dari berbagai pihak. Pada kesempatan ini penulis menyampaikan terima kasih kepada :

1. Allah SWT karena berkatnya penyusun diberi kemudahan untuk menyelesaikan skripsi ini tepat waktu.
2. Kedua orang tua saya, serta semua keluarga yang selalu mendoakan, memberikan dukungan dan semangat dalam menjalani kuliah dan menyelesaikan skripsi ini.
3. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta
4. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta
5. Ibu Windha Mega Pradnya D, M.Kom selaku ketua program studi S1 Informatika Universitas AMIKOM Yogyakarta
6. Bapak Nuri Cahyono, M.Kom. sebagai dosen pembimbing skripsi yang telah memberikan bimbingan dan ilmu baru kepada penulis.
7. Bapak dan Ibu Dosen FIK AMIKOM Yogyakarta.
8. Seluruh civitas akademika Universitas AMIKOM Yogyakarta yang telah memberikan pengetahuan dan pengalaman kepada penulis selama perkuliahan.

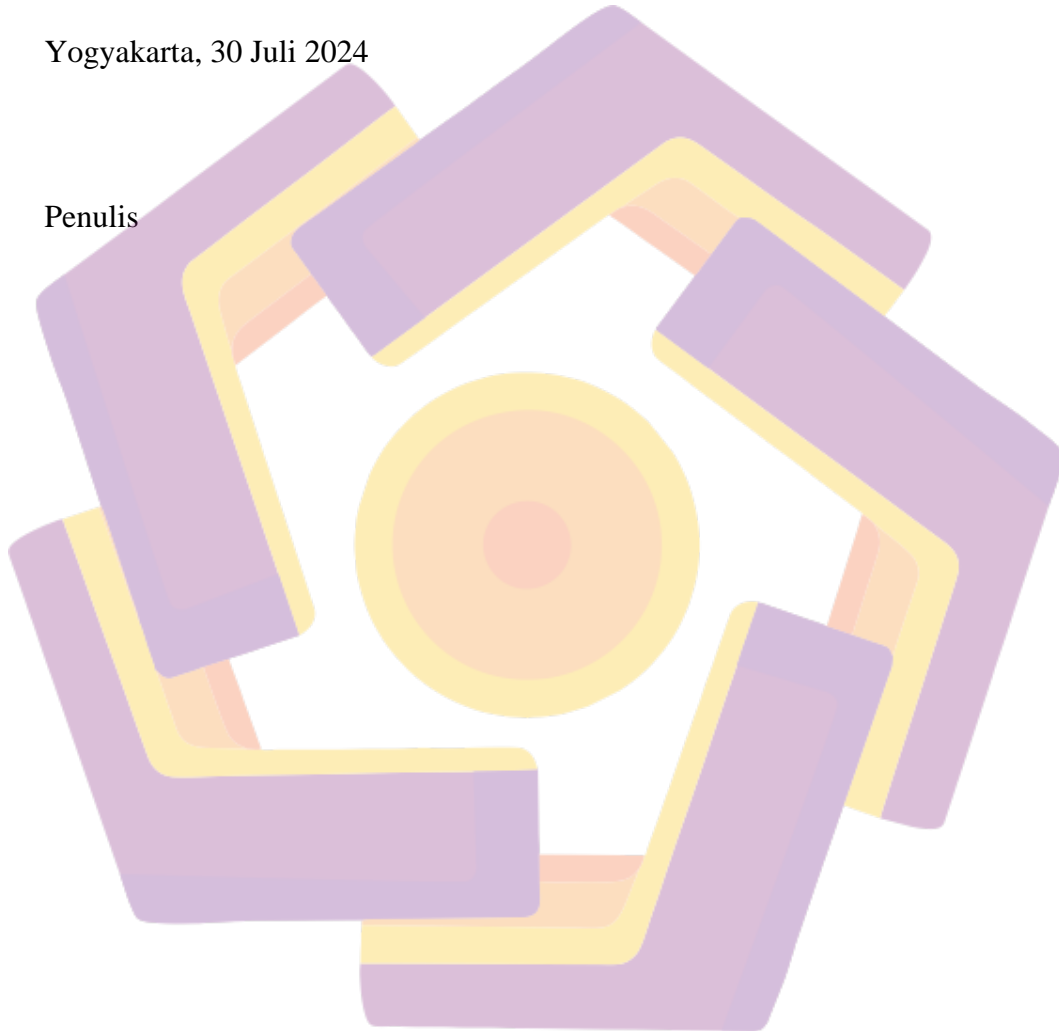
Penyusun menyadari bahwa laporan ini masih jauh dari kata kesempurnaan. Oleh karena itu kritik dan saran yang membangun untuk perbaikan laporan ini sangat penulis harapkan dari semua pihak.

Wassalamu 'alaikum Wr . Wb

Semoga skripsi ini dapat bermanfaat bagi penulis, pembaca dan pihak-pihak lainnya.

Yogyakarta, 30 Juli 2024

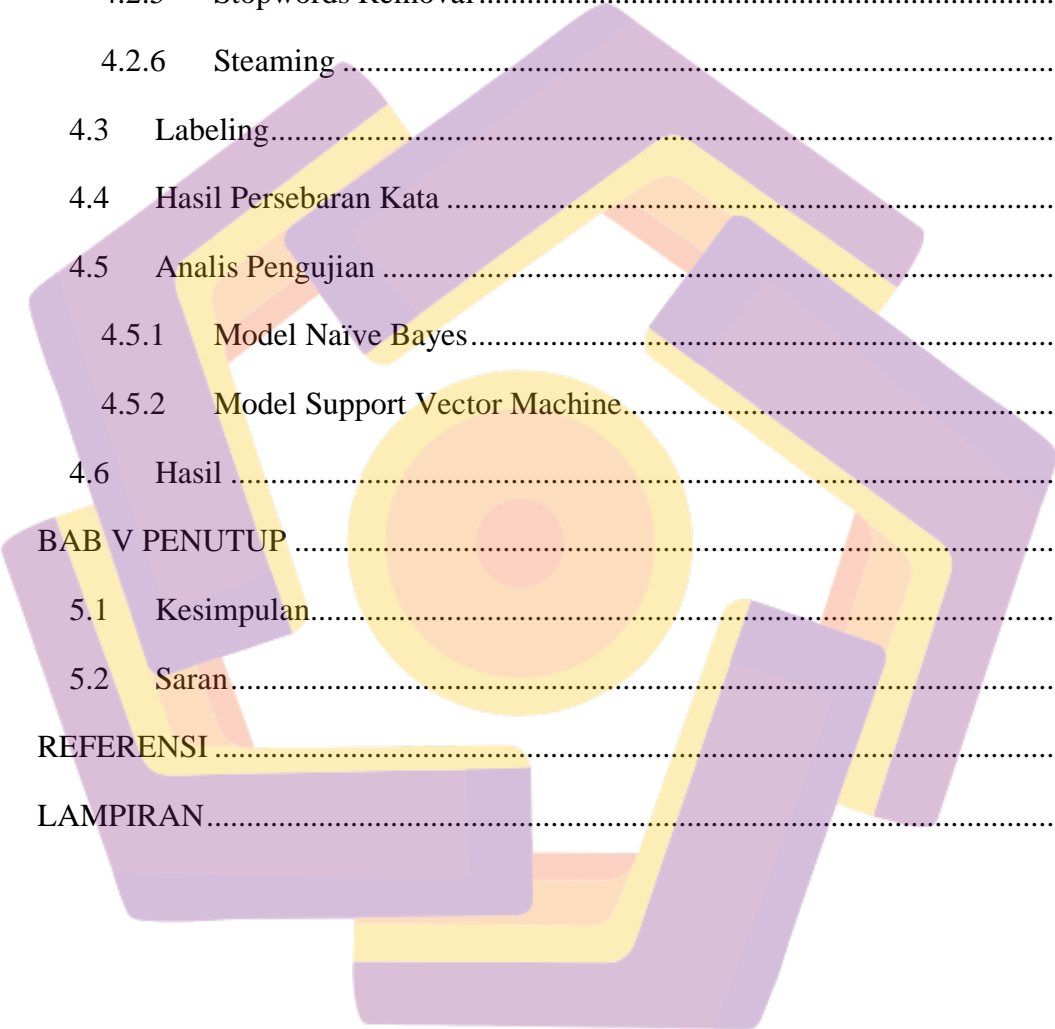
Penulis



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiv
DAFTAR LAMBANG DAN SINGKATAN	xv
DAFTAR ISTILAH	xvi
INTISARI	xvii
<i>ABSTRACT</i>	xix
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.....	10

2.2.1	Machine Learning	10
2.2.2	Text Mining	10
2.2.3	Fungsi Text Mining.....	12
2.2.4	Klasifikasi Data Mining.....	12
2.2.5	Analisis Sentimen	13
2.2.6	Youtube.....	14
2.2.7	Lexicon Classification	14
2.2.8	CountVecorizer	15
2.2.9	WordCloud.....	15
2.2.10	Naive Bayes	16
2.2.11	Support Vector Machine (SVM).....	17
2.2.12	Performance Evaluation Measure (Confusion Matrix).....	18
BAB III METODE PENELITIAN		20
3.1	Objek Penelitian	20
3.2	Alur penelitian.....	20
3.3	Metode Pengolahan Data.....	21
3.3.1	Pengumpulan Data	21
3.3.2	Preprocessing	23
3.3.3	Labelling	30
3.3.4	Klasifikasi	31
3.3.5	Confusion Matrix	32
3.4	Alat dan Bahan	32
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Pengambilan data	34
4.2	Pre-processing Data.....	35

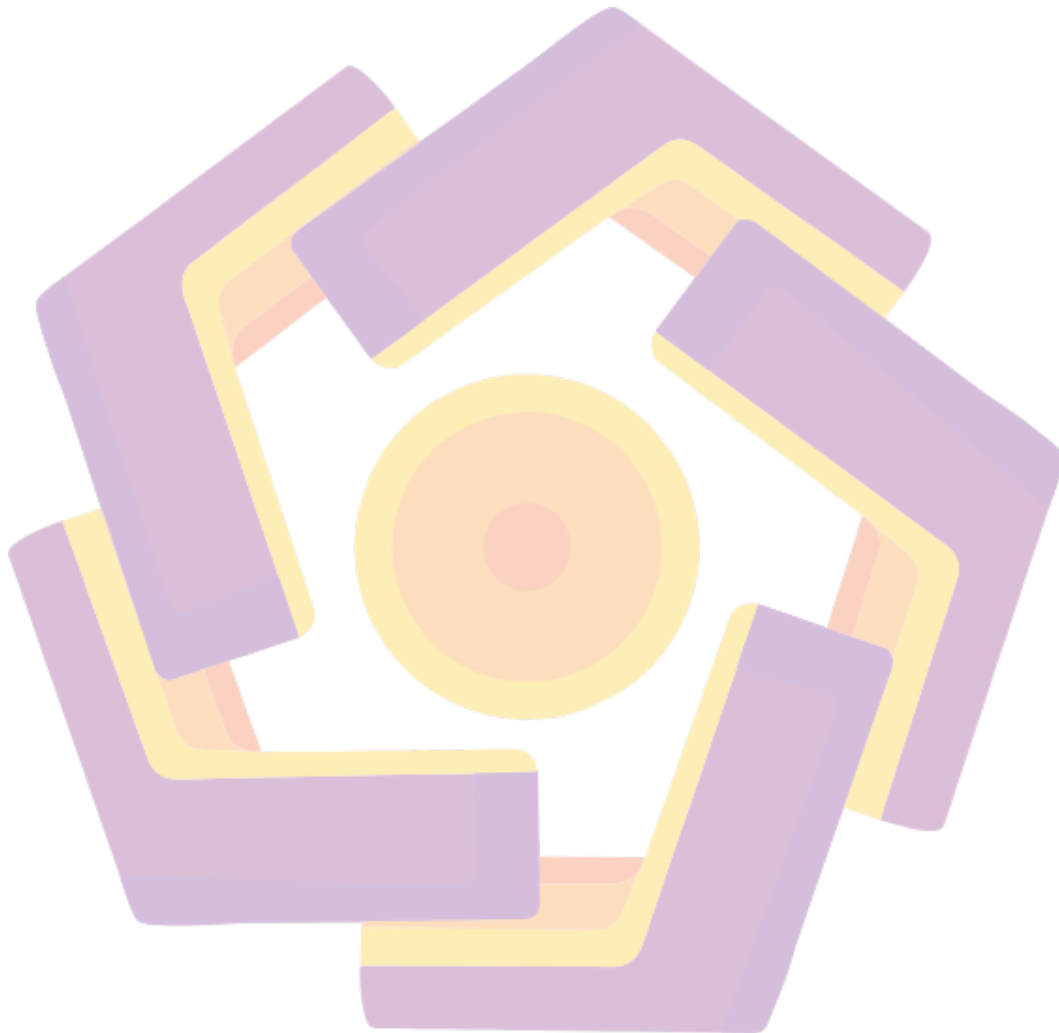


4.2.1	Case Folding	35
4.2.2	Cleaning	36
4.2.3	Normalisasi	37
4.2.4	Tokenization	38
4.2.5	Stopwords Removal	39
4.2.6	Steaming	40
4.3	Labeling.....	41
4.4	Hasil Persebaran Kata	43
4.5	Analisis Pengujian	44
4.5.1	Model Naïve Bayes.....	45
4.5.2	Model Support Vector Machine.....	49
4.6	Hasil	53
BAB V PENUTUP		55
5.1	Kesimpulan.....	55
5.2	Saran.....	55
REFERENSI		56
LAMPIRAN.....		60

DAFTAR TABEL

Tabel 2. 1. Keaslian Penelitian.....	7
Tabel 2. 2 Keaslian Penelitian.....	8
Tabel 2. 3 Lanjutan Keaslian Penelitian	9
Tabel 2. 4 Confusion Matrix	18
Tabel 3. 1 Case Folding	24
Tabel 3. 2 Contoh tanda baca, simbol dan angka.....	25
Tabel 3. 3 Contoh Cleaning	26
Tabel 3. 4 Contoh Normalisasi	27
Tabel 3. 5 Contoh Tokenization.....	28
Tabel 3. 6 Contoh Tokenization.....	29
Tabel 3. 7 Contoh Stemming	30
Tabel 3. 8 Alat Penelitian.....	33
Tabel 4. 1 Hasil Case Folding.....	36
Tabel 4. 2 Hasil Cleaning.....	37
Tabel 4. 3 Hasil Normaliasi	38
Tabel 4. 4 Hasil Normaliasi	39
Tabel 4. 5 Hasil Stopword Removal	40
Tabel 4. 6 Hasil Steaming.....	41
Tabel 4. 7 Hasil Labeling.....	42
Tabel 4. 8 Persebaran data	42
Tabel 4. 9 Hasil Evaluasi Complement naïve bayes.....	45
Tabel 4. 10 Confusion Matrix Complement Naïve Bayes.....	46
Tabel 4. 11 Tabel Pengujian performa Naïve bayes Complement	47
Tabel 4. 12 Hasil Evaluasi Gaussion	47
Tabel 4. 13 Confusion Matrix Gaussion	48
Tabel 4. 14 Hasil pengujian performa Gaussion NB	49
Tabel 4. 15 Hasil Kernel Linier	50
Tabel 4. 16 Confusion Matrix Kernel Linier	50

Tabel 4. 17 Hasil Pengujian Performa Kernel Linier.....	51
Tabel 4. 18 Hasil Kernel RBF.....	52
Tabel 4. 19 Confusion Matrix Kernel RBF	52
Tabel 4. 20 Hasil Pengujian Performa Kernel RBF.....	53
Tabel 4. 21 Hasil akurasi keseluruhan	54

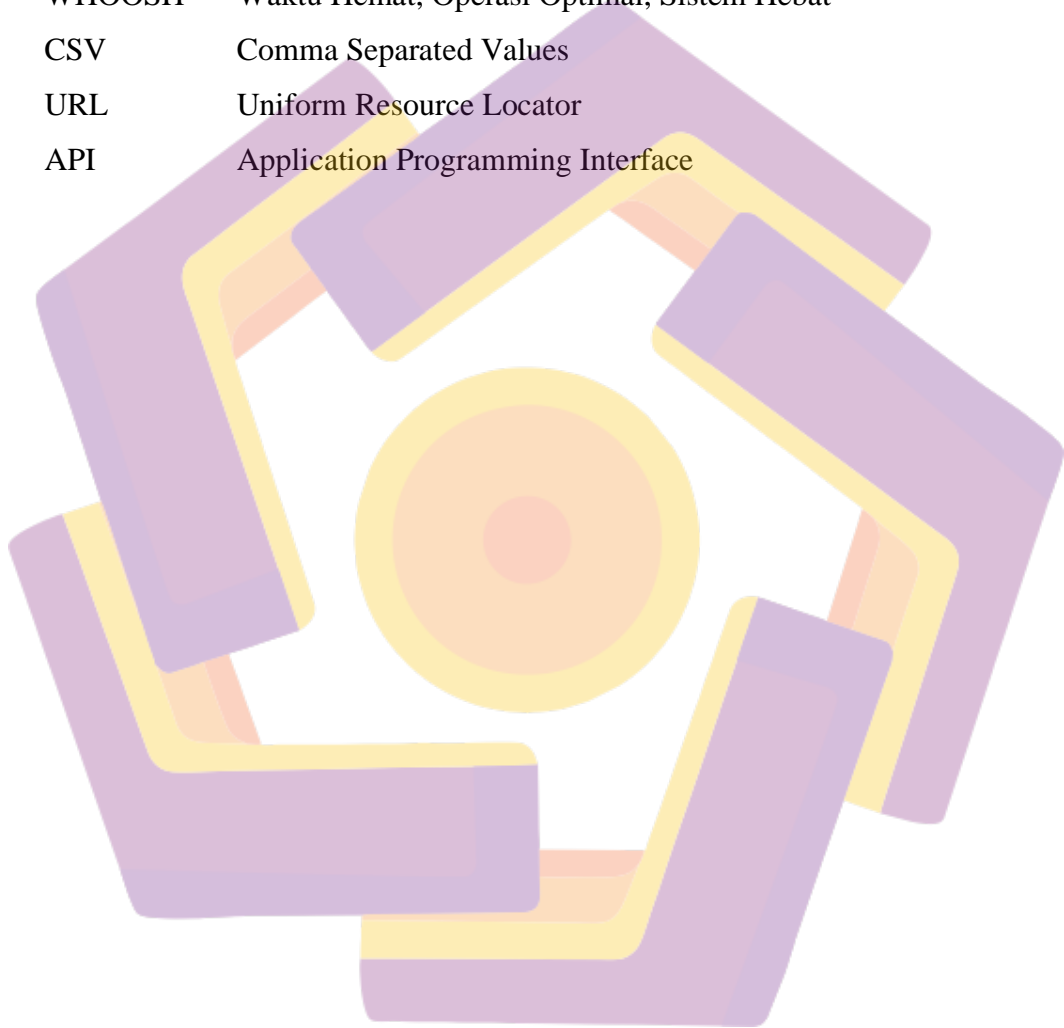


DAFTAR GAMBAR

Gambar 2. 1 Tahap Text Mining.....	12
Gambar 3. 1 Alur Penelitian	20
Gambar 3. 2 Alur Pengumpulan Data.....	21
Gambar 3. 3 codingan pengambilan data pada Google Colab.....	22
Gambar 3. 4 Data komentar Youtube CNN Indonesia mengenai Kereta Cepat Jakarta-Bandung	23
Gambar 3. 5 Alur Pre-processing.....	24
Gambar 3. 6 Alur Case Folding	24
Gambar 3. 7 Alur Cleaning.....	25
Gambar 3. 8 Alur Normalisasi	26
Gambar 3. 9 Alur Tokenization	27
Gambar 3. 10 Alur Stopword Removal	28
Gambar 3. 11 Alur Steaming	30
Gambar 4. 1 Dataset.....	34
Gambar 4. 2 Tampilan dataset yang hanya diambil isi komentarnya	35
Gambar 4. 3 Source Code Case Folding.....	35
Gambar 4. 4 Source code Cleaning.....	36
Gambar 4. 5 Source Code Normaliasi	37
Gambar 4. 6 Source Code Tokenisasi.....	38
Gambar 4. 7 Source Code Stopwords Removal.....	39
Gambar 4. 8 Source Code Steaming	40
Gambar 4. 9 Source Code Lexicon Based	41
Gambar 4. 10 Pie Chart pembagian Sentiment Compound Score	43
Gambar 4. 11 Persebaran kata Positif	44
Gambar 4. 12 Persebaran kata Nertral	44
Gambar 4. 13 Persebaran kata Negatif.....	44

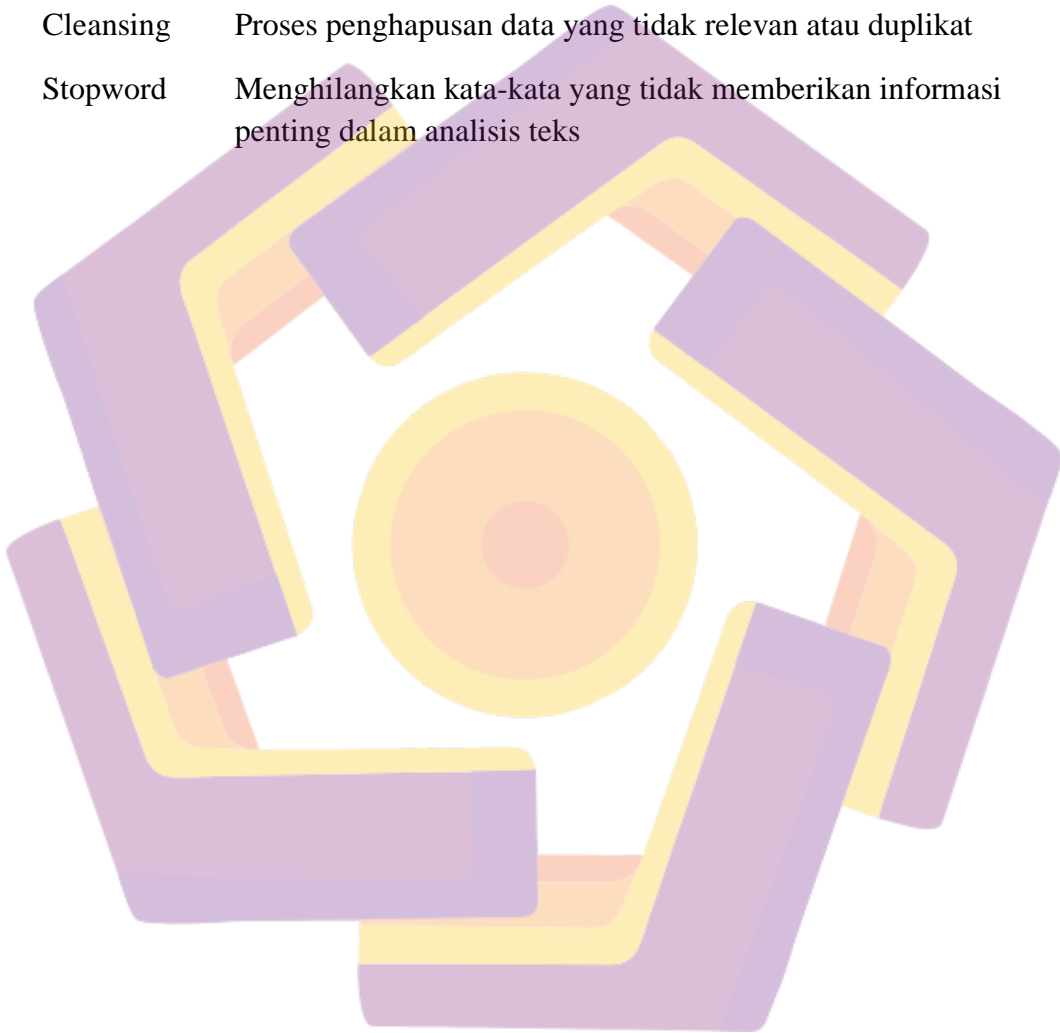
DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
NB	Naïve Bayes
VADER	Valence Aware Dictionary and Sentiment Reasoner
WHOOSH	Waktu Hemat, Operasi Optimal, Sistem Hebat
CSV	Comma Separated Values
URL	Uniform Resource Locator
API	Application Programming Interface



DAFTAR ISTILAH

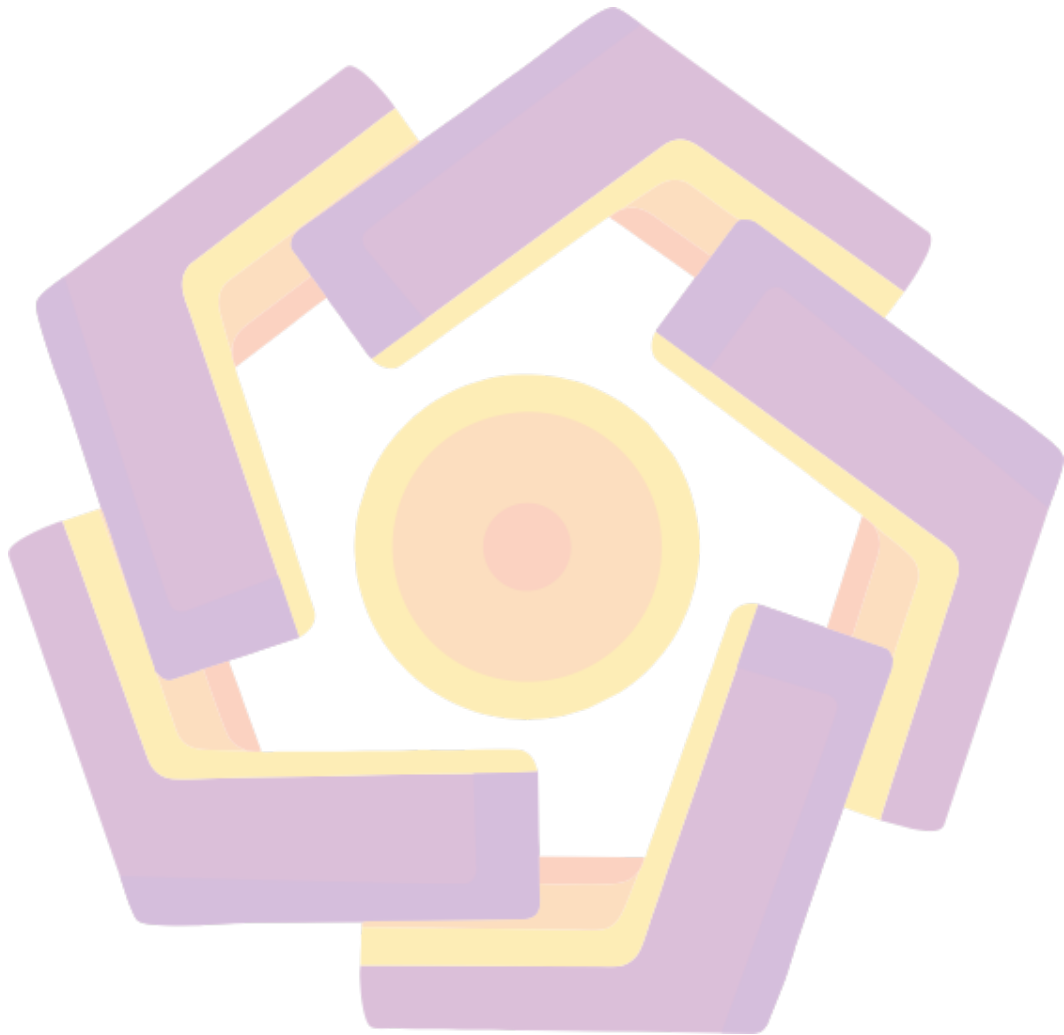
Case Folding	Proses mengubah semua data teks menjadi huruf kecil (lowercase)
Normalize	Tahap mengubah kata-kata tidak baku menjadi baku
Tokenize	Memisahkan kata menjadi kalimat terpisah
Cleansing	Proses penghapusan data yang tidak relevan atau duplikat
Stopword	Menghilangkan kata-kata yang tidak memberikan informasi penting dalam analisis teks



INTISARI

Penelitian ini bertujuan untuk membandingkan kinerja dua metode klasifikasi, yaitu Support Vector Machine (SVM) dan Naïve Bayes, dalam analisis sentimen komentar YouTube pada kanal CNN Indonesia terkait proyek Kereta Cepat Jakarta-Bandung (WHOOSH). Analisis sentimen merupakan proses otomatisasi yang digunakan untuk mengidentifikasi dan mengelompokkan opini atau emosi dari teks, yang dalam hal ini berasal dari komentar pengguna di platform YouTube. Data yang digunakan dalam penelitian ini terdiri dari komentar yang dikumpulkan dari video terkait topik Kereta Cepat Jakarta-Bandung di kanal YouTube CNN Indonesia. Proses penelitian ini meliputi beberapa tahapan utama, yaitu pengumpulan data, preprocessing, serta penerapan metode klasifikasi SVM dan Naïve Bayes. Kinerja kedua metode kemudian dibandingkan berdasarkan beberapa metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa masing-masing metode memiliki kelebihan dan kekurangan tersendiri dalam menganalisis sentimen. Hasil pengujian dan evaluasi Support Vector Machine (SVM) yaitu Kernel Linier dan Kernel RBF. Untuk Kernel Linier mendapatkan 81,3%, Presisi 80,9%, Recall 79,5%, F1-Score 80%. Sedangkan untuk Kernel RBF mendapatkan akurasi sebesar 80,9%, Presisi 81,6%, Recall, 78,8%, F1-Score 80%. Lalu untuk hasil pengujian Naïve Bayes yang dibagi menjadi dua yaitu Complement Naïve Bayes dan Gaussion NB. Untuk Complement Naïve Bayes mendapatkan akurasi sebesar 63,98%, Presisi 63,3%, Recall 62,6% dan F-1 Score 61%. Sedangkan untk Gaussion NB mendapatkan 64,41%, Presisi 62,4%, Recall 62,1%, F1-Score 62%. Hasil penelitian ini dapat menunjukkan bahwa Support Vector Machine (SVM) performa lebih baik daripada Naïve bayes. Temuan ini memberikan wawasan berharga bagi peneliti dan praktisi dalam memilih metode klasifikasi yang paling sesuai untuk analisis sentimen komentar di media sosial, khususnya pada platform YouTube.

Kata kunci: Sentiment Analysis, Support Vector Machine, Naïve Bayes, YouTube Comments, Jakarta-Bandung High-Speed Rail, WHOOSH



ABSTRACT

This study aims to compare the performance of two classification methods, Support Vector Machine (SVM) and Naïve Bayes, in sentiment analysis of YouTube comments on CNN Indonesia's channel regarding the Jakarta-Bandung High-Speed Rail (WHOOSH) project. Sentiment analysis is an automated process used to identify and classify opinions or emotions from text, which in this case comes from user comments on the YouTube platform. The data used in this study consists of comments collected from videos related to the Jakarta-Bandung High-Speed Rail topic on CNN Indonesia's YouTube channel. The research process involves several main stages: data collection, preprocessing, and the application of SVM and Naïve Bayes classification methods. The performance of both methods is then compared based on several evaluation metrics such as accuracy, precision, recall, and F1-score. The results of the study indicate that each method has its own advantages and disadvantages in analyzing sentiment. The testing and evaluation results for Support Vector Machine (SVM) with Linear Kernel showed an accuracy of 81.3%, precision of 80.9%, recall of 79.5%, and F1-score of 80%. For the RBF Kernel, the accuracy was 80.9%, precision was 81.6%, recall was 78.8%, and F1-score was 80%. For Naïve Bayes, the results are divided into Complement Naïve Bayes and Gaussian NB. Complement Naïve Bayes achieved an accuracy of 63.98%, precision of 63.3%, recall of 62.6%, and F1-score of 61%. Gaussian NB achieved an accuracy of 64.41%, precision of 62.4%, recall of 62.1%, and F1-score of 62%. The findings of this study demonstrate that Support Vector Machine (SVM) performs better than Naïve Bayes. These findings provide valuable insights for researchers and practitioners in selecting the most appropriate classification method for sentiment analysis of social media comments, particularly on the YouTube platform.

Keyword: Sentiment Analysis, Support Vector Machine, Naïve Bayes, YouTube Comments, Jakarta-Bandung High-Speed Rail, WHOOSH

