

**SENTIMEN ANALISIS MENGGUNAKAN TEKNOLOGI
MACHINE LEARNING DENGAN MEMBANDINGKAN ALGORITMA
KLASIFIKASI RF DAN SVM PADA TWEET ADUAN
TERHADAP PT GARUDA INDONESIA**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

ANDIKA MAHENDRA IRAWAN

20.11.3460

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

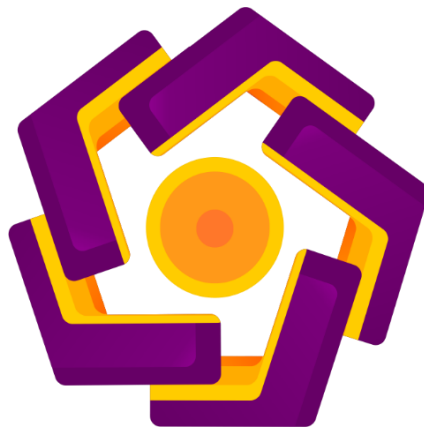
YOGYAKARTA

2024

**SENTIMEN ANALISIS MENGGUNAKAN TEKNOLOGI
MACHINE LEARNING DENGAN MEMBANDINGKAN ALGORITMA
KLASIFIKASI RF DAN SVM PADA TWEET ADUAN
TERHADAP PT GARUDA INDONESIA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

ANDIKA MAHENDRA IRAWAN

20.11.3460

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS SAMIKOMYOGYAKARTA**

YOGYAKARTA

2024

HALAMAN PERSETUJUAN

SKRIPSI

**SENTIMEN ANALISIS MENGGUNAKAN TEKNOLOGI
MACHINE LEARNING DENGAN MEMBANDINGKAN ALGORITMA
KLASIFIKASI RF DAN SVM PADA TWEET ADUAN
TERHADAP PT GARUDA INDONESIA**

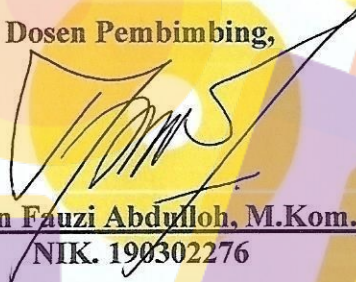
yang disusun dan diajukan oleh

Andika Mahendra Irawan

20.11.3460

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 8 Mei 2024

Dosen Pembimbing,



Ferian Fauzi Abdulloh, M.Kom.
NIK. 190302276

HALAMAN PENGESAHAN
SKRIPSI
SENTIMEN ANALISIS MENGGUNAKAN TEKNOLOGI
MACHINE LEARNING DENGAN MEMBANDINGKAN ALGORITMA
KLASIFIKASI RF DAN SVM PADA TWEET ADUAN
TERHADAP PT GARUDA INDONESIA

yang disusun dan diajukan oleh

Andika Mahendra Irawan

20.11.3460

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Mei 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Sharazita Dyah Anggita, M,Kom
NIK. 190302285

Ike Verawati, M,Kom
NIK. 190302237

Ferian Fauzi Abdulloh, M,Kom
NIK. 190302276



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Mei 2024

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Andika Mahendra Irawan**
NIM : **20.11.3460**

Menyatakan bahwa Skripsi dengan judul berikut:

SENTIMEN ANALISIS MENGGUNAKAN TEKNOLOGI MACHINE LEARNING DENGAN MEMBANDINGKAN ALGORITMA KLASIFIKASI RF DAN SVM PADA TWEET ADUAN TERHADAP PT GARUDA INDONESIA

Dosen Pembimbing : **Ferian Fauzi Abdulloh, 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 Mei 2024

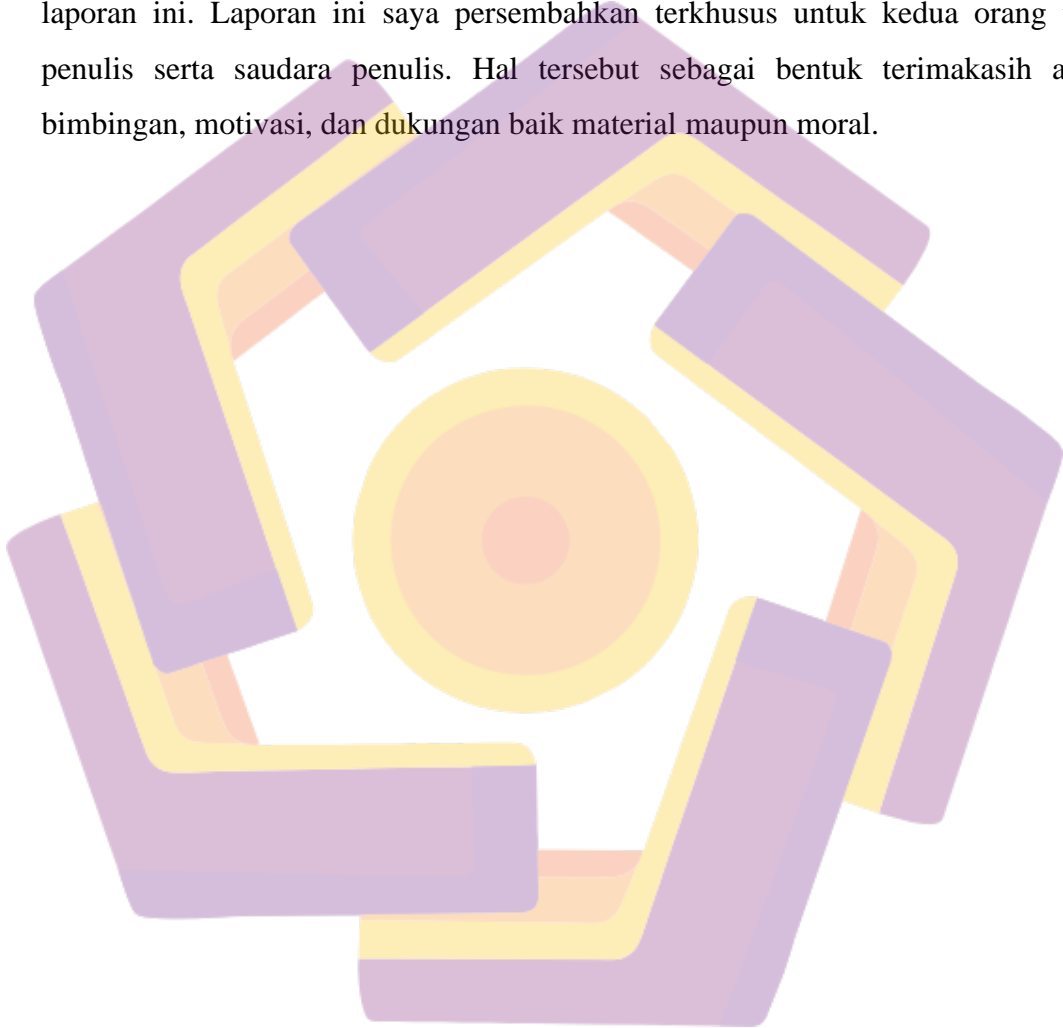
Yang Menyatakan,



Andika Mahendra Irawan

HALAMAN PERSEMBAHAN

Puji dan Syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, dengan berkat dan Rahmat-Nya penulisan diberikan Kesehatan untuk menyelesaikan laporan ini. Laporan ini saya persembahkan terkhusus untuk kedua orang tua penulis serta saudara penulis. Hal tersebut sebagai bentuk terimakasih atas bimbingan, motivasi, dan dukungan baik material maupun moral.



KATA PENGANTAR

Penyusunan laporan tugas akhir ini sebagai syarat dalam menyelesaikan studi jenjang Sarjana (S1) Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Ucapan terima kasih penulis sampaikan kepada pihak yang mendukung dan berkontribusi penulis dalam menyelesaikan penelitian ini, antara lain Ferian Fauzi Abdulloh, M.Kom, Kedua orang tua serta saudara penulis yang telah memberikan dukungan baik material maupun moral, Teman-teman Angkatan 2020 yang telah banyak memberikan masukan dalam segala bentuk diskusi dan menyelesaikan penelitian tugas akhir.

Yogyakarta, 20 Mei 2024



Andika Mahendra Irawan

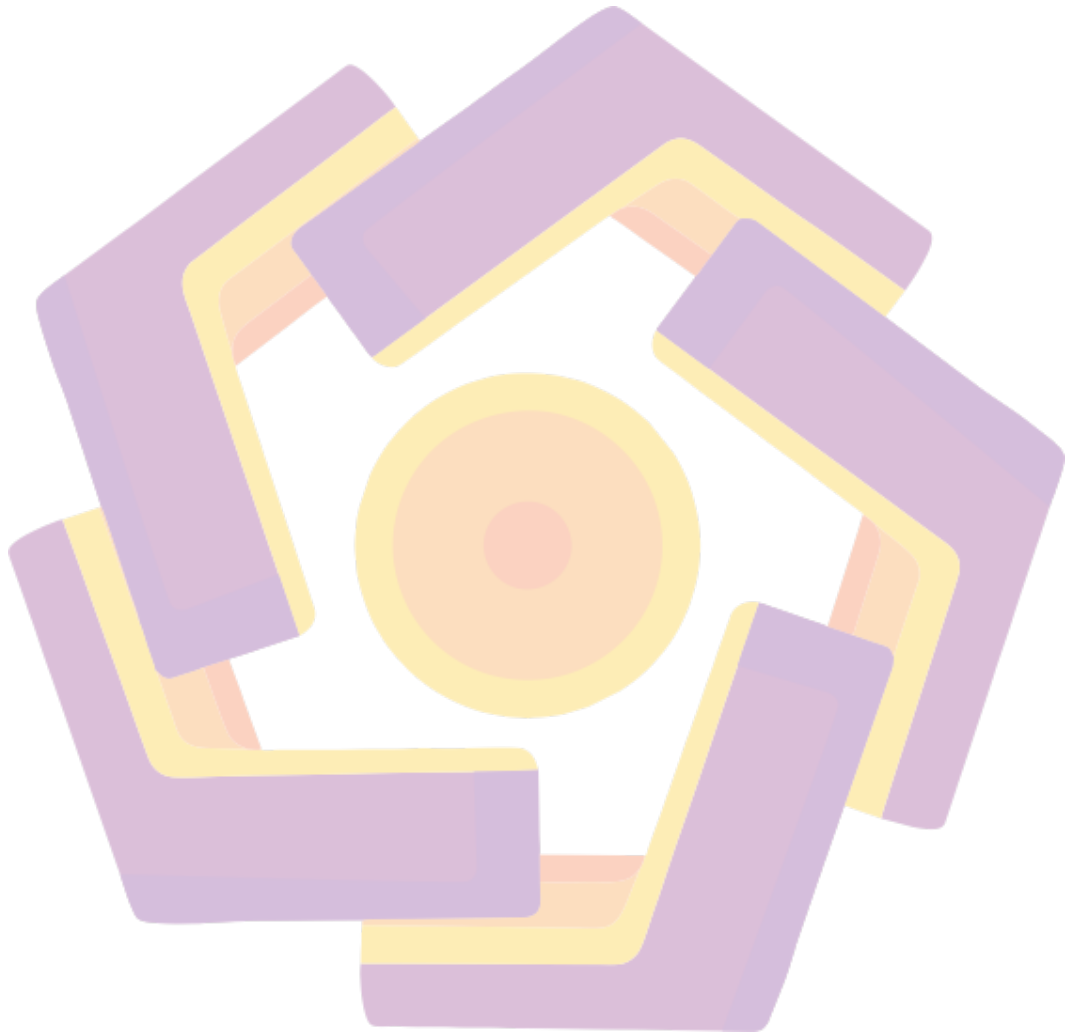
DAFTAR ISI

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	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	11
2.2.1 Media sosial dan Twitter.....	11
2.2.2 Data Preprocessing.....	13
2.2.3 Sentimen Analysis.....	13
2.2.4 Data Mining	15

2.2.4.1	Fungsi data mining	15
2.2.5	Machine Learning	17
2.2.6	Algoritma Klasifikasi	17
2.2.7	Random Forest	18
2.1.8	Support Vector Machine	21
2.1.9	Confusion Matrix	24
2.1.10	Split Validation	25
2.1.11	Cross Validation	25
BAB III METODE PENELITIAN		27
3.1	Objek Penelitian	27
3.2	Alur Penelitian	27
3.2.1	Dataset Penelitian	27
3.2.1.1	Import Dataset	28
3.2.2	Data Preprocessing	28
3.2.2.1	Lower Casing	28
3.2.2.2	Removal of Punctuation	29
3.2.2.3	Removal of Stopwords	29
3.2.2.4	Removal of URLs and other Unnecessary Character	29
3.2.2.5	Removal of Emojis and Emoticons	29
3.2.2.6	Removal of Frequent Words	30
3.2.2.7	Stemming	30
3.2.2.8	Stemming Bahasa Indonesia	30
3.2.2.9	Chat Words Conversion	31
3.2.2.10	Bag of Words	31
3.2.2.11	TF-IDF	32
3.2.3	DataModeling	33
3.2.3.1	Support Vector Machine	33
3.2.3.2	Random Forest	34
3.2.4	Data Evaluation	35
3.2.4.1	Confusion Matrix	35

3.2.4.2	Split Validation	37
3.2.4.3	Cross-Validation	38
3.3	Alat dan Bahan.....	38
3.3.1.	Alat.....	38
3.3.2.	Bahan	39
BAB IV HASIL DAN PEMBAHASAN		40
4.1	Upload Dataset.....	40
4.2	Text Analytic.....	40
4.2.1	Get Top n-Words	41
4.2.2	Create Wordcloud	43
4.3	Preprocessing	44
4.3.1	Lower Casing	44
4.3.2	Removal of Punctuation.....	45
4.3.3	Removal of Stopwords.....	46
4.3.4	Removal of URLs and other Unnecessary Character	47
4.3.5	Removal of Emojis and Emoticons.....	47
4.3.6	Removal of Frequent Words	49
4.3.7	Stemming	50
4.3.8	Stemming Bahasa Indonesia	51
4.3.9	Chat Words Conversion	52
4.3.10	Identifying The Final Result	54
4.4	Feature Extraction	56
4.4.1	Bag of Words	57
4.4.2	TF-IDF	57
4.5	Train Model.....	58
4.5.1	Support Vector Machine	58
4.5.2	Random Forest Classifier.....	59
4.6	Model Validasi.....	60
4.6.1	Split Validation	61

4.6.2	Cross-Validation	63
4.7	Hasil Experimen.....	65
BAB V PENUTUP		66
5.1	Kesimpulan	66
5.2	Saran	66
REFERENSI		67



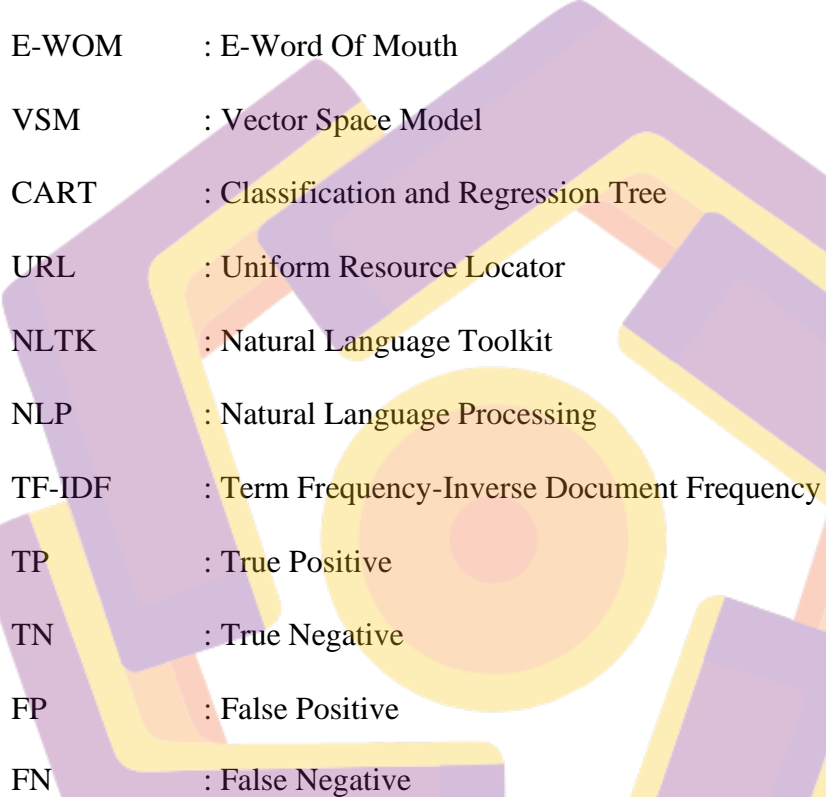
DAFTAR TABEL

TABEL 2.1. KEASLIAN PENELITIAN	9
TABEL 2.2. CONFUSION MATRIX	24
TABEL 4.1. KATEGORI LABEL DATA	40
TABEL 4.2. TWEET SEBELUM DAN SESUDAH LOWER CASING	44
TABEL 4.3. TWEET SEBELUM DAN SESUDAH PENGHAPUSAN TANDA BACA	45
TABEL 4.4. TWEET SEBELUM DAN SESUDAH PENGHAPUSAN STOPWORDS	46
TABEL 4.5. TWEET SEBELUM DAN SESUDAH PENGHAPUSAN EMOJIS	48
TABEL 4.6. TWEET SEBELUM DAN SESUDAH PENGHAPUSAN FREQUENT WORDS	50
TABEL 4.7. TWEET SEBELUM DAN SESUDAH STEMMING	51
TABEL 4.8. TWEET SEBELUM DAN SESUDAH STEMMING BAHASA INDONESIA.....	52
TABEL 4.9. CHAT WORDS CONVERSION.....	53
TABEL 4.10. PERBANDINGAN HASIL SPLIT VALIDATION.....	61
TABEL 4.11. CROSS-VALIDATION RESULT.....	63

DAFTAR GAMBAR

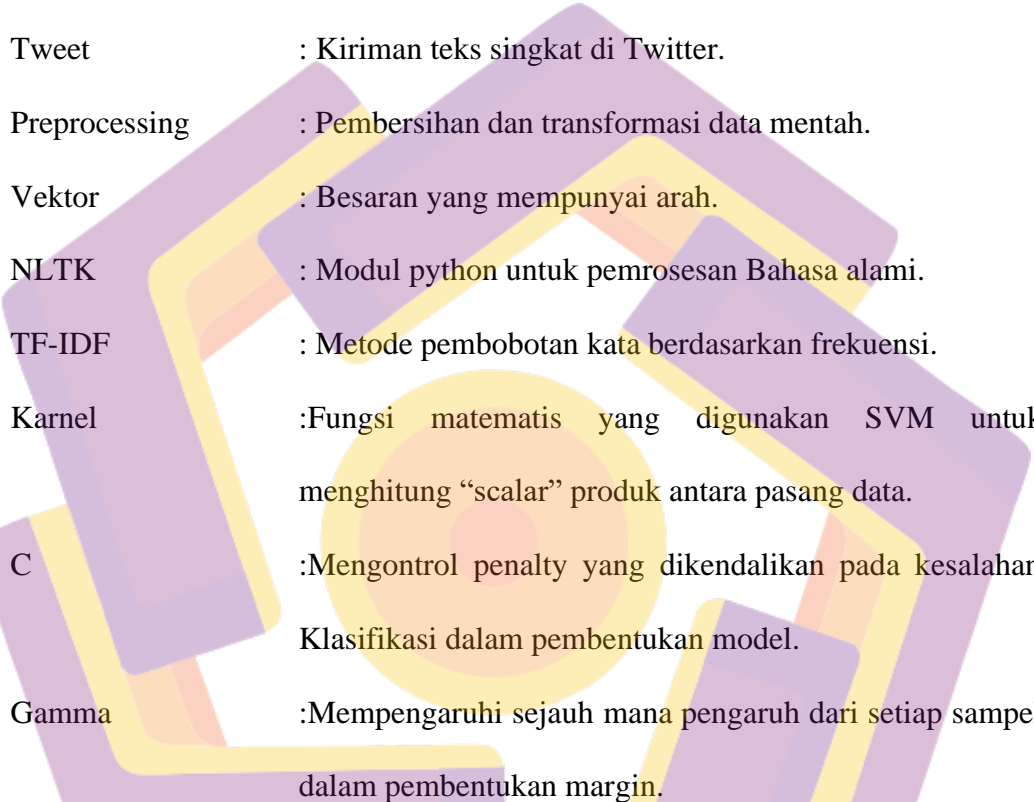
GAMBAR 2.1. DIAGRAM ASPEK DASAR PENGGUNAAN MEDIA SOSIAL	11
GAMBAR 2.2. TAHAPAN KLASIFIKASI TEKS	14
GAMBAR 2.3. METODE DATA MINING	16
GAMBAR 2.4. DECISION TREE	18
GAMBAR 2.5. PENGAMBILAN KEPUTUSAN RF	20
GAMBAR 2.6. GARIS DISKRIMINASI.....	21
GAMBAR 2.7. HYPERPLANE SVM	22
GAMBAR 2.8. HYPERPLANE DAN MARGIN	23
GAMBAR 2.9. SKEMA CROSS VALIDATION.....	26
GAMBAR 3.1. ALUR PENELITIAN	27
GAMBAR 4.1. DATASET PENELITIAN.....	40
GAMBAR 4.2. DISTRIBUSI LABEL DATA	41
GAMBAR 4.3. KATA-KATA YANG SERING MUNCUL PADA HALAMAN ATAS	42
GAMBAR 4.4. WORDCLOUT SEBELUM DILAKUKAN PREPROCESSING.....	43
GAMBAR 4.5. TWEET YANG SERING MUNCUL SECARA BERULANG PADA DATASET.....	49
GAMBAR 4.6. HASIL AKHIR PREPROCESSING DATA.....	55
GAMBAR 4.7. WORDCLOUT SESUDAH DILAKUKAN PREPROCESSING.....	56

DAFTAR LAMBANG DAN SINGKATAN



RF	: Random Forest
SVM	: Support Vector Machines
BOW	: Bag Of Words
E-WOM	: E-Word Of Mouth
VSM	: Vector Space Model
CART	: Classification and Regression Tree
URL	: Uniform Resource Locator
NLTK	: Natural Language Toolkit
NLP	: Natural Language Processing
TF-IDF	: Term Frequency-Inverse Document Frequency
TP	: True Positive
TN	: True Negative
FP	: False Positive
FN	: False Negative

DAFTAR ISTILAH



Dataset	: Sekumpulan data.
Analisis sentiment	: Pengenalan dan ekstraksi opini dari teks.
Twitter	: Media sosial berbasis tweet.
Tweet	: Kiriman teks singkat di Twitter.
Preprocessing	: Pembersihan dan transformasi data mentah.
Vektor	: Besaran yang mempunyai arah.
NLTK	: Modul python untuk pemrosesan Bahasa alami.
TF-IDF	: Metode pembobotan kata berdasarkan frekuensi.
Karnel	: Fungsi matematis yang digunakan SVM untuk menghitung “scalar” produk antara pasang data.
C	: Mengontrol penalty yang dikendalikan pada kesalahan Klasifikasi dalam pembentukan model.
Gamma	: Mempengaruhi sejauh mana pengaruh dari setiap sampel dalam pembentukan margin.

INTISARI

Pengembangan Model Sentimen Analisis dengan Teknologi Machine Learning pada Data Tweet Aduan Terhadap Public service di bidang penerbangan. Konektivitas melalui media sosial di zaman sekarang telah memungkinkan interaksi luas antara manusia. Penelitian ini bertujuan untuk mengembangkan model sentimen analisis menggunakan teknologi machine learning untuk menganalisis tweet aduan terhadap Public service di bidang penerbangan. Penelitian ini membandingkan dua metode klasifikasi Random Forest dan Support Vector Machine dalam analisis sentimen dan mengidentifikasi topik yang paling berpengaruh dalam data media sosial perusahaan. Data tweet aduan dikumpulkan melalui API Twitter dengan kata kunci terkait perusahaan dan hanya tweet aduan yang publik yang digunakan. Tujuannya adalah untuk menemukan metode klasifikasi dengan akurasi terbaik dalam mendeteksi sentimen dan mengidentifikasi topik utama. Penelitian ini memberikan sumbangan bagi penelitian penerapan teknik penambangan data pada media sosial serta pemahaman lebih lanjut tentang sentimen dan opini publik terhadap merek perusahaan.

Kata kunci: Machine Learning, SVM, Random Forest, Sentiment Analysis.

ABSTRACT

Development of Sentiment Analysis Model with Machine Learning Technology on Complaint Tweets Data Towards Public Service in the Aviation Sector. Connectivity through social media in today's era has enabled extensive interaction among humans. This research aims to develop a sentiment analysis model using machine learning technology to analyze complaint tweets towards Public service in the aviation sector. The study compares two classification methods, Random Forest and Support Vector Machine, in sentiment analysis and identifies the most influential topics in company social media data. Complaint tweet data was collected via the Twitter API with keywords related to the company, and only public complaint tweets were used. The goal is to find the classification method with the best accuracy in detecting sentiment and identifying main topics. This research contributes to the application of data mining techniques on social media and a deeper understanding of sentiment and public opinion towards company brands.

Keywords: Machine Learning, SVM, Random Forest, Sentiment Analysis.