

KOMPREHENSIF

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
KEBIJAKAN KENAIAKAN PPN 12% MENGGUNAKAN
METODE *SUPPORT VECTOR MACHINE* DAN *NAIVE BAYES***



SANDI PRATAMA

1512618046

PROGRAM STUDI

PENDIDIKAN TEKNIK INFORMATIKA DAN KOMPUTER

FAKULTAS TEKNIK

UNIVERSITAS NEGERI JAKARTA

2025

LEMBAR PENGESAHAN KOMPREHENSIF

Judul : ANALISIS SENTIMEN PENGGUNA TWITTER
TERHADAP KEBIJAKAN KENAIKAN PPN 12%
MENGUNAKAN METODE *SUPPORT VECTOR
MACHINE* DAN *NAIVE BAYES*

Penyusun : Sandi Pratama

NIM : 1512618046

Tanggal Ujian : 23 Juli 2025

Disetujui oleh:

Pembimbing I,



Dr. Widodo, S.Kom, M.Kom

NIP. 197203252005011002

Pengesahan Panitia Ujian Skripsi :

Ketua Penguji,



Ressa Dwitias Sari, S.T., M.T.I

NIP. 198909152019032021

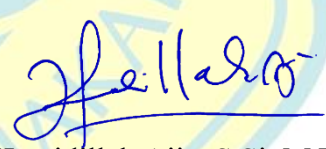
Anggota Penguji I,



Nur Elah, S.Kom., M.T., CSA

NIP. 199104042024062002

Anggota Penguji II,



Hamidillah Ajie, S.Si, M.T

NIP. 197408242005011001

Mengetahui,

Koordinator Program Studi Pendidikan Teknik Informatika dan Komputer

Intelligentia - Dignitas

Muhammad Ficky Duskarnaen, M.Sc

NIP. 197309242006041001

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Komprehensif ini merupakan Karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Komprehensif ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta

Jakarta, 27 Juni 2022

Yang membuat pernyataan



Sandi Pratama

No. Reg. 1512618046

Intelligentia - Dignitas



KEMENTERIAN PENDIDIKAN TINGGI, SAINS DAN TEKNOLOGI
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220

Telepon/Faksimili: 021-4894221

Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Sandi Pratama

NIM : 1512618046

Fakultas/Prodi : Fakultas Teknik/Pendidikan Teknik Informatika dan Komputer

Alamat email : sansan00668@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☐ Skripsi ☐ Tesis ☐ Disertasi ☒ Lain-lain (Komprehensif)

yang berjudul :

Analisis Sentimen Pengguna Twitter Terhadap Kebijakan Kenaikan PPN 12% Menggunakan Metode Support Vector Machine dan Naive Bayes

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 29 Juli 2025

Penulis

(Sandi Pratama)

nama dan tanda tangan

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan penelitian ini dengan judul “*Analisis Sentimen Pengguna Twitter terhadap Kenaikan PPN 12% Menggunakan SVM dan Naïve Bayes*” sebagai salah satu syarat untuk menyelesaikan tugas akhir.

Shalawat dan salam senantiasa tercurah kepada junjungan Nabi Muhammad Shallallahu ‘alaihi wa sallam, keluarga, sahabat, serta seluruh umatnya yang istiqamah mengikuti ajarannya hingga akhir zaman.

Penelitian ini dilakukan untuk menganalisis respons masyarakat terhadap kebijakan kenaikan Pajak Pertambahan Nilai (PPN) melalui media sosial Twitter. Penulis menggunakan pendekatan analisis sentimen dengan algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes* guna mengklasifikasikan opini menjadi positif, netral, dan negatif. Penulis berharap hasil dari penelitian ini dapat memberikan kontribusi dalam pengembangan ilmu di bidang *text mining* dan analisis data.

Dalam kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Orang tua tercinta, atas doa yang tiada henti, kepercayaan, dan dukungan moril yang tak tergantikan. Terima kasih karena selalu yakin bahwa penulis mampu menyelesaikan studi ini, meski harus melalui proses yang cukup panjang.
2. Ketua Program Studi, Bapak Muhammad Ficky Duskarnaen, M.Sc., atas bimbingan, perhatian, dan dorongan yang terus-menerus diberikan kepada mahasiswa, serta kesediaannya membantu dalam hal administratif dan kendala selama proses penyusunan skripsi.
3. Dosen pembimbing, Dr. Widodo, S.Kom., M.Kom., yang telah dengan sabar dan penuh ketulusan membimbing penulis. Terima kasih atas segala masukan, arahan, dan waktu yang telah diberikan hingga penelitian ini dapat diselesaikan dengan baik.
4. Teman-teman, yang selalu hadir memberi semangat, mengingatkan di saat mulai menyerah, serta menjadi tempat berbagi keluh kesah selama

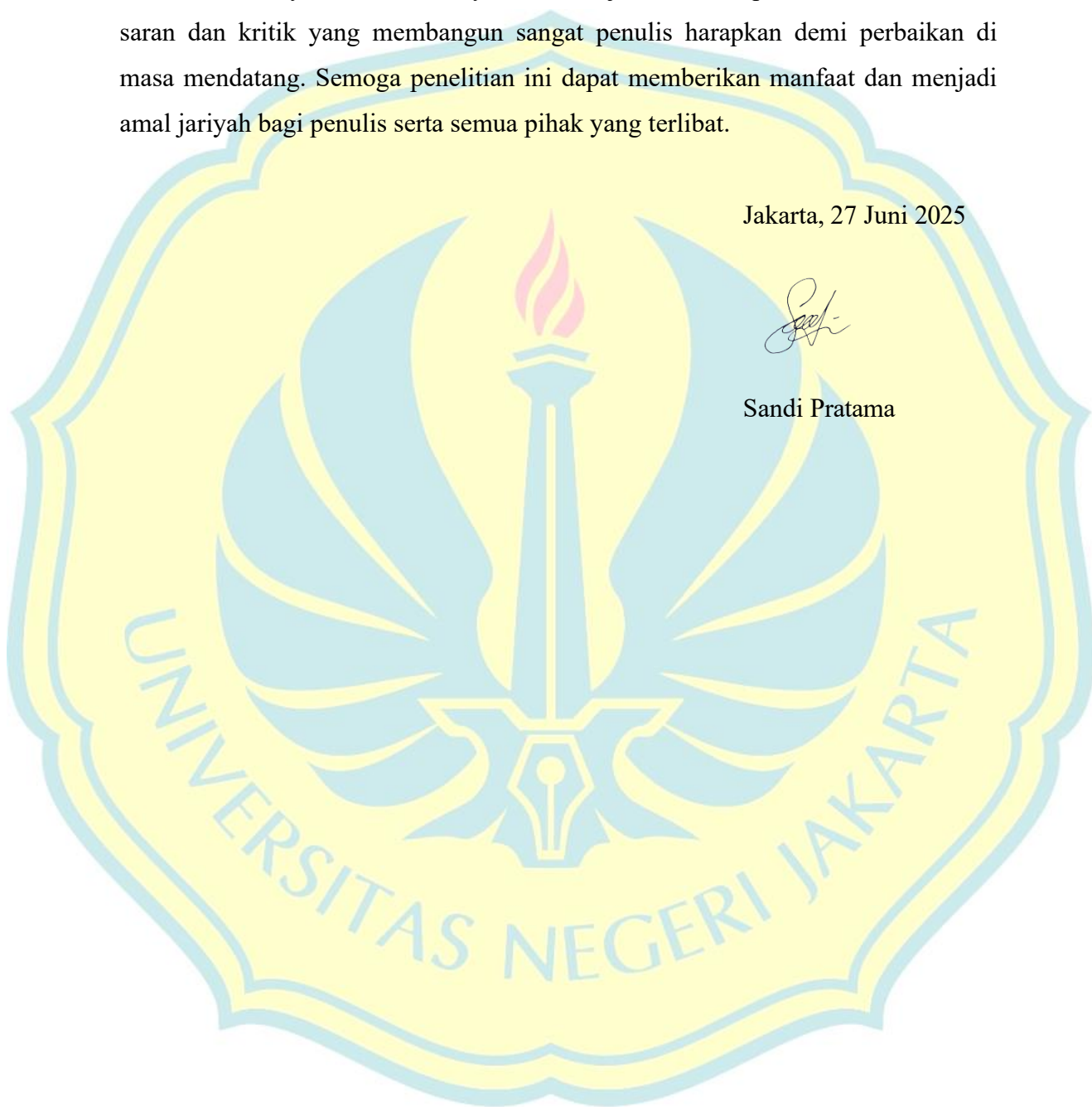
proses menyusun skripsi ini. Terima kasih atas dorongan moral dan keceriaan yang membuat perjalanan ini terasa lebih ringan.

Penulis menyadari bahwa karya ini masih jauh dari sempurna. Oleh karena itu, saran dan kritik yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Semoga penelitian ini dapat memberikan manfaat dan menjadi amal jariyah bagi penulis serta semua pihak yang terlibat.

Jakarta, 27 Juni 2025



Sandi Pratama



Intelligentia - Dignitas

ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP KEBIJAKAN KENAIKAN PPN 12% MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* DAN *NAIVE BAYES*

Sandi Pratama

Dosen Pembimbing : Dr. Widodo, S.Kom, M.Kom

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen pengguna Twitter terhadap kebijakan kenaikan Pajak Pertambahan Nilai (PPN) sebesar 12% di Indonesia. Data dikumpulkan melalui proses *web crawling* terhadap *tweet* berbahasa Indonesia yang relevan dengan topik selama periode November 2024 hingga Januari 2025. Proses analisis dilakukan dengan pendekatan *text mining* yang meliputi *Preprocessing* teks, pembobotan menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)*, dan klasifikasi sentimen menggunakan dua algoritma *machine learning* yaitu *Naïve Bayes* dan *Support Vector Machine (SVM)*. Sentimen diklasifikasikan ke dalam tiga kategori: positif, netral, dan negatif. Hasil evaluasi menunjukkan bahwa SVM memiliki performa lebih tinggi dibandingkan *Naïve Bayes* dengan akurasi sebesar 88,2% dibandingkan 74,6%. Mayoritas *tweet* yang dianalisis menunjukkan sentimen netral, meskipun terdapat porsi signifikan sentimen negatif yang mencerminkan kekhawatiran masyarakat terhadap kebijakan tersebut. Penelitian ini memberikan kontribusi dalam pemanfaatan analisis media sosial sebagai alat evaluasi respons publik terhadap kebijakan PPN.

Kata kunci: Analisis sentimen, Twitter, PPN 12%, *Support Vector Machine*, *Naïve Bayes*, TF-IDF

Intelligentia - Dignitas

SENTIMEN ANALYSIS OF TWITTER USERS TOWARD THE 12% VAT INCREASE POLICY USING *SUPPORT VECTOR MACHINE* AND *NAÏVE BAYES* METHODS

Sandi Pratama

Advisor: Dr. Widodo, S.Kom, M.Kom

ABSTRACT

This study aims to analyze the sentiment of Twitter users regarding the policy of increasing Indonesia's Value Added Tax (VAT) to 12%. Data were collected through a web *crawling* process targeting Indonesian-language *tweets* relevant to the topic, spanning the period from November 2024 to January 2025. The analysis was conducted using a *text mining* approach, which included *text Preprocessing*, weighting using *Term Frequency-Inverse Document Frequency (TF-IDF)*, and sentiment classification using two *machine learning* algorithms: *Naïve Bayes* and *Support Vector Machine (SVM)*. Sentiments were categorized into three classes: positive, neutral, and negatif. Evaluation results indicated that the SVM algorithm outperformed *Naïve Bayes* with an *accuracy* of 88.2% compared to 74.6%. Most of the analyzed *tweets* expressed a neutral sentiment, although a significant portion showed negatif sentiment, reflecting public concern over the policy. This study contributes to the utilization of social media analysis as a tool for evaluating public response to *fiscal* policy.

Keywords: Sentimen analysis, Twitter, VAT 12%, *Support Vector Machine*, *Naïve Bayes*, TF-IDF

Intelligentia - Dignitas

DAFTAR ISI

LEMBAR PENGESAHAN KOMPREHENSIF	ii
LEMBAR PERNYATAAN.....	iii
LEMBAR PERNYATAAN PUBLIKASI	iv
KATA PENGANTAR	v
ABSTRAK.....	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
 BAB I.....	 1
PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Identifikasi Masalah	2
1.3. Batasan Masalah.....	3
1.4. Rumusan Masalah	3
1.5. Tujuan Penelitian.....	3
1.6. Manfaat Penelitian.....	3
 BAB II.....	 5
LANDASAN TEORI.....	5
2.1 Landasan Teori	5
2.1.1 Pajak Pertambahan Nilai (PPN)	5
2.1.2 Twitter	5
2.1.3 <i>Text Mining</i>	6
2.1.4 Analisis Sentimen.....	7
2.1.5 <i>Text Preprocessing</i>	7
2.1.5.1 <i>Cleansing</i>	8
2.1.5.2 <i>Case folding</i>	8
2.1.5.3 <i>Tokenization</i>	9

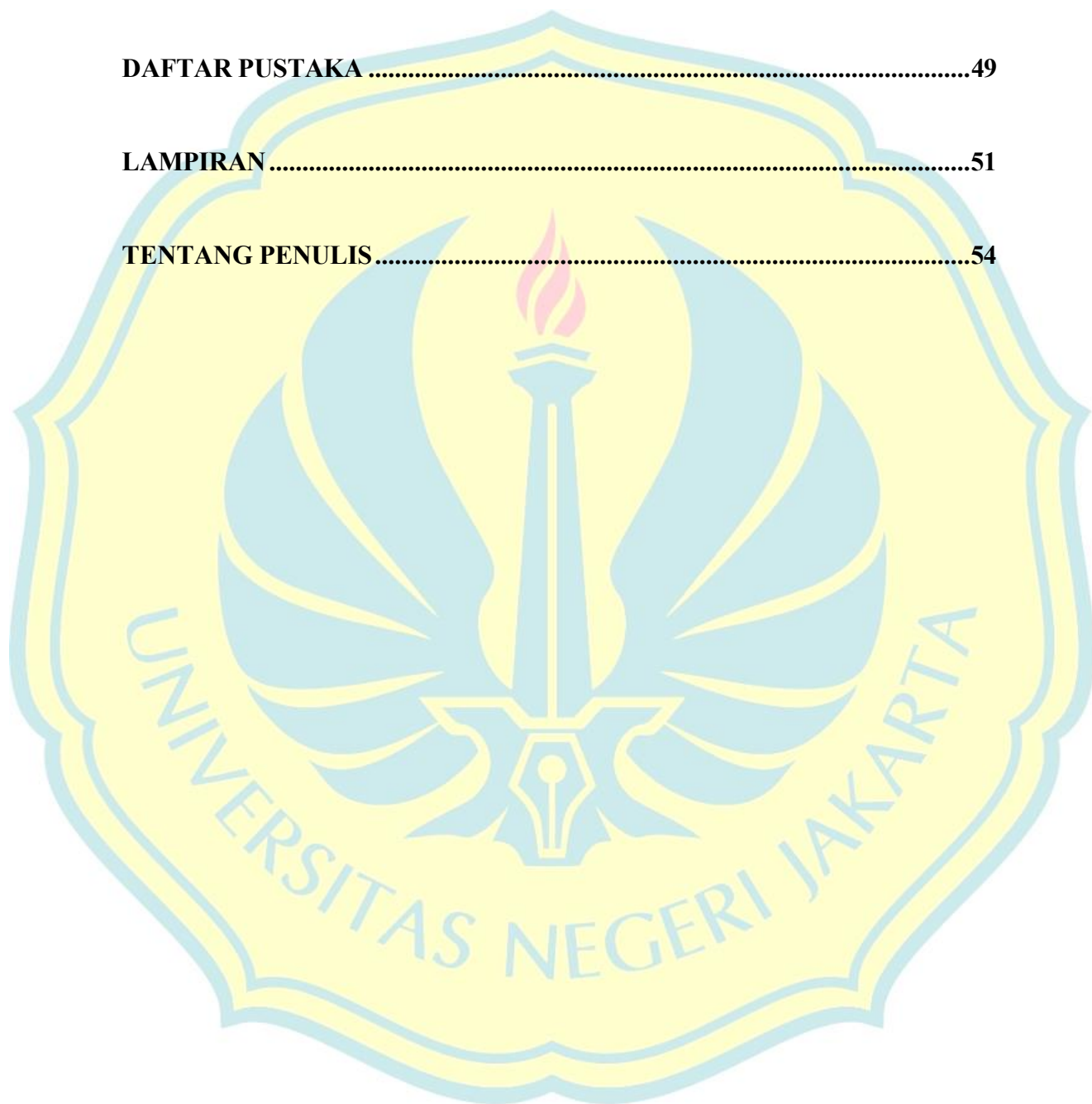
2.1.5.4	<i>Filtering</i>	9
2.1.5.5	<i>Normalization</i>	10
2.1.5.6	<i>Stemming</i>	10
2.1.5.7	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	11
2.1.6	<i>Support Vector Machine (SVM)</i>	12
2.1.7	<i>Naïve Bayes (NB)</i>	13
2.1.8	Evaluasi Kinerja	14
2.1.8.1	<i>Precision</i>	15
2.1.8.2	<i>Recall</i>	15
2.1.8.3	<i>F1-score</i>	16
2.1.8.4	<i>Accuracy</i>	16
2.2	Penelitian Relevan.....	17
2.3	Kerangka Berfikir.....	26
BAB III		28
HASIL PENELITIAN.....		28
3.1	Pengumpulan Data.....	28
3.2	<i>Preprocessing</i>	30
3.2.1	<i>Cleansing</i> dan <i>Case folding</i>	31
3.2.2	<i>Tokenization</i>	32
3.2.3	Stopword Removal.....	34
3.2.4	Normalisasi.....	35
3.2.5	<i>Stemming</i>	37
3.2.6	Penghapusan Duplikasi	38
3.3	Pembobotan TF-IDF.....	39
3.4	Klasifikasi Sentimen.....	41
3.5	Pengklasifikasian dengan Model SVM dan <i>Naïve Bayes</i>	42
3.5.1	Proses Klasifikasi Menggunakan <i>Naïve Bayes</i>	42
3.5.2	Proses Klasifikasi Menggunakan <i>Support Vector Machine (SVM)</i>	43
3.6	Evaluasi Kinerja dan Distribusi Hasil	44
BAB IV		47
KESIMPULAN DAN SARAN		47

4.1	Kesimpulan.....	47
4.2	Saran.....	48

DAFTAR PUSTAKA	49
-----------------------------	-----------

LAMPIRAN	51
-----------------------	-----------

TENTANG PENULIS	54
------------------------------	-----------



Intelligentia - Dignitas

DAFTAR TABEL

Tabel 2.1 Contoh Hasil <i>Cleanshing</i>	8
Tabel 2.2 Hasil <i>Case folding</i>	8
Tabel 2.3 Hasil Tokenisasi	9
Tabel 2.4 Hasil Dari <i>Filtering</i>	9
Tabel 2.5 Contoh Hasil Normalisasi.....	10
Tabel 2.6 Contoh Hasil <i>Stemming</i>	11
Tabel 2.7 Daftar Penelitian Relevan.....	20
Tabel 3.1 Daftar Kata Positif dan Negatif.....	42
Tabel 3.2 Hasil Klasifikasi Naive Bayes dan SVM	44
Tabel 3.3 Hasil Klasifikasi Preprocessing dan TF-IDF.....	46



Intelligentia - Dignitas

DAFTAR GAMBAR

Gambar 2.1 Visualiasi SVM.....	13
Gambar 2.2 Kerangka Berfikir	27
Gambar 3.1 Pengambilan auth_token Twitter.....	29
Gambar 3.4 Hasil Tokenisasi	33
Gambar 3.5 Hasil <i>Filtering</i>	35
Gambar 3.6 Hasil Normalisasi	37
Gambar 3.7 Hasil <i>Stemming</i>	38
Gambar 3.8 Hasil Akurasi Naive Bayes.....	43
Gambar 3.9 Hasil Akurasi SVM	44
Gambar 3.10 Hasil <i>Confusion matrix</i> Naive Bayes.....	45
Gambar 3.11 Hasil <i>Confusion matrix</i> SVM	45



Intelligentia - Dignitas

DAFTAR LAMPIRAN

Lampiran 1 – Data Hasil <i>Crawling</i>	50
Lampiran 2 – Data Hasil <i>Cleansing</i>	50
Lampiran 3 – Data <i>Case folding</i>	50
Lampiran 4 – Data Tokenisasi	50
Lampiran 5 – Data Setelah Filtered	50
Lampiran 6 – Kamus Normalisasi	50
Lampiran 7 – Data Setelah Normalisasi	51
Lampiran 8 – Kamus <i>Stemming</i>	51
Lampiran 9 – Data Setelah <i>Stemming</i>	51
Lampiran 10 – Data Tanpa Duplikat	51
Lampiran 11 – Hasil TF-IDF	51
Lampiran 12 – Data Label Manual	51
Lampiran 13 – Data Hasil Klasifikasi Otomatis	51
Lampiran 14 – Diagram Batang Distribusi Sentimen	52
Lampiran 15 – Diagram Lingkaran Proporsi Sentimen	52
Lampiran 16 – <i>Confusion matrix: Naïve Bayes</i>	52
Lampiran 17 – <i>Confusion matrix: SVM</i>	52

Intelligentia - Dignitas