

TUGAS AKHIR

**ANALISIS SENTIMEN TERHADAP ANIES BASWEDAN
MENGUNAKAN METODE SUPPORT VECTOR MACHINE
STUDI KASUS MEDIA SOSIAL TWITTER**

Disusun Untuk Memenuhi Persyaratan Meraih Gelar Sarjana Komputer Teknik
Informatika Muhammadiyah Jember



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2023

HALAMAN PERSETUJUAN

ANALISIS SENTIMEN TERHADAP ANIES BASWEDAN
MENGUNAKAN METODE SUPPORT VECTOR MACHINE
STUDI KASUS MEDIA SOSIAL TWITTER

Oleh:

Bagas Farasqa Nauval Akbar

1710651095

Disetujui oleh:

Dosen Pembimbing I

Dosen Pembimbing II


Deni Arifianto, M.Kom

NIDN. 0718068103


Amalina Maryam Zakiyyah, SE., M.Si

NIDN. 0727058705

HALAMAN PENGESAHAN

ANALISIS SENTIMEN TERHADAP ANIES BASWEDAN
MENGUNAKAN METODE SUPPORT VECTOR MACHINE
STUDI KASUS MEDIA SOSIAL TWITTER

Oleh:

Bagas Farasqa Nauval Akbar

1710651095

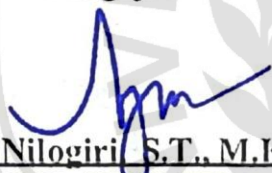
Telah diuji dan dipertahankan pada,

Hari, Tanggal : Selasa, 08 Agustus 2023

Tempat : Ruang Rapat Dosen TI

Disetujui oleh:

Dosen Penguji:
Penguji I



Agung Nilogiri, S.T., M.Kom
NIDN. 0030037701

Dosen Pembimbing:
Pembimbing I



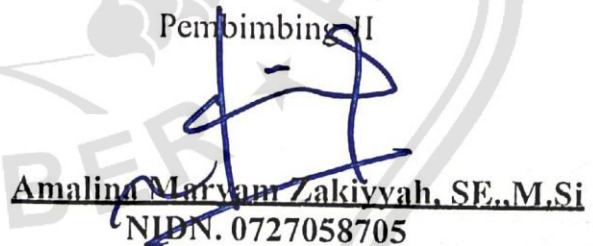
Deni Arifianto, M.Kom
NIDN. 0718068103

Penguji II



Querota A'yun, M.Pd
NIDN. 0703069002

Pembimbing II



Amalina Maryani Zakriyyah, SE., M.Si
NIDN. 0727058705

Mengesahkan,
Dekan Fakultas Teknik



Dr. Nanang Saiful Rizal, S.T., MT., IPM
NIDN. 0705047806

Mengetahui,
Ketua Program Studi Teknik
Informatika



Adi Eko Wardoyo, S.T., M.Kom
NIDN. 0014027501

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Bagas Farasqa Nauval Kbar
NIM : 1710651095
Institusi : S1 Teknik Informatika, Fakultas Teknik.
Universitas Muhammadiyah Jember

Menyatakan bahwa Tugas Akhir yang berjudul “Analisis Sentimen Terhadap Anies Baswedan Menggunakan Metode *Support Vector Machine* Studi Kasus Media Sosial Twitter” bukan merupakan Tugas Akhir milik orang lain baik sebagian maupun keseluruhan kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak benar penulis bersedia mendapatkan sanksi dari akademik.

Jember, 08 Agustus 2023

Yang menyatakan:



Bagas Farasqa Nauval Akbar

NIM. 1710651095



KATA PENGANTAR

Segala puji dan syukur kami panjatkan kehadirat Allah SWT, tuhan yangmaha kuasa karena berkat rahmat dan inayah-nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Analisis Sentimen Terhadap Anies Baswedan Menggunakan Metode *Support Vector Machine* Studi Kasus Media Sosial Twitter”. Dalam penyusunan tugas akhir ini penulis menghadapi banyak hambatan serta rintangan namun pada akhirnya berkat bantuan dari berbagai pihak penulis mampu menyelesaikan tugas akhir ini. Untuk itu penulis menyampaikan terimakasih kepada:

1. Bapak Deni Arifianto, M.Kom dan Ibu Amalina Maryam Zakiyyah, SE., M.Si selaku pembimbing yang telah memberikan bimbingan, ilmu dan arahan.
2. Bapak Agung Nilogiri, S.T., M.Kom dan Ibu Qurrota A'yun, M.Pd selaku penguji yang telah memberikan kritik dan saran.
3. Bapak Dr. Nanang Saiful Rizal, ST., MT., IPM selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
4. Bapak Ari Eko Wardoyo, S.T., M.Kom selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Jember.
5. Seluruh dosen, staf, dan karyawan Universitas Muhammadiyah Jember atas jasa-jasanya selama penulis menuntut ilmu.
6. Orang tua saya, keluarga, serta teman-teman yang telah memberikan dukungan berupa material dan moral.

Penulis menyadari bahwa dalam menyusun laporan ini terdapat kekurangan dan keterbatasan, oleh karena itu kritik dan saran sangat diharapkan. Akhir kata penulis berharap semoga laporan Tugas Akhir ini dapat bermanfaat bagi pengembangan ilmu pengetahuan.

Jember, 08 Agustus 2023

Penulis

MOTTO

“Boleh jadi kamu membenci sesuatu padahal ia amat baik bagimu, dan boleh jadi pula kamu menyukai sesuatu padahal ia amat buruk bagimu, Allah mengetahui sedang kamu tidak mengetahui”

(QS. Al-Baqarah : 216)



**ANALISIS SENTIMEN TERHADAP ANIES BASWEDAN
MENGUNAKAN METODE SUPPORT VECTOR MACHINE STUDI
KASUS MEDIA SOSIAL TWITTER**

ABSTRAK

Anies Baswedan menjadi salah satu bakal calon presiden yang digulirkan media massa menjelang digelarnya Pemilihan Presiden 2024-2029 mendatang mulai menjadi topik pemberitaan media massa di Indonesia. Salah satu media sosial yang sering dipakai oleh publik yaitu Twitter. Banyaknya pengguna media sosial di Indonesia membuat media sosial terutama Twitter dapat digunakan secara efektif untuk melihat bagaimana pendapat dari masyarakat. Dari berbagai opini dan pendapat di twitter dibutuhkan sebuah teknik untuk membagi opini ke dalam kelas opini negatif, netral atau positif. Pada penelitian ini, digunakan Support Vector Machine Multiclass untuk proses klasifikasi dengan menggunakan metode One Against Rest dan Oversampling untuk mengatasi ketidakseimbangan kelas. Hasil dari penelitian ini mendapatkan kesimpulan bahwa masyarakat cenderung memiliki sentimen negatif terhadap Anies Baswedan. Hasil dari pengujian klasifikasi Support Vector Machine tanpa proses balancing data didapatkan nilai Accuracy sebesar 77%, Precision 76% dan Recall 72%, setelah proses balancing data didapatkan nilai Accuracy sebesar 97%, Precision 83% dan Recall 94%.

Kata Kunci: Analisis sentimen, Twitter, SVM, Oversampling, One Against Rest.

**SENTIMENT ANALYSIS OF ANIES BASWEDAN USING THE SUPPORT
VECTOR MACHINE METHOD CASE STUDY OF TWITTER SOCIAL
MEDIA**

ABSTRACT

Anies Baswedan is one of the presidential candidates being rolled out by the mass media ahead of the upcoming 2024-2029 Presidential Election, starting to become the topic of mass media coverage in Indonesia. One of the social media that is often used by the public is Twitter. The large number of social media users in Indonesia means that social media, especially Twitter, can be used effectively to see what the public thinks. From the various opinions and opinions on Twitter, a technique is needed to divide opinions into negative, neutral or positive opinion classes. In this study, a Multiclass Support Vector Machine is used for the classification process using the One Against Rest and Oversampling method to overcome class imbalance. The results of this study conclude that people tend to have negative sentiments towards Anies Baswedan. The results of the Support Vector Machine classification test without the data balancing process obtained Accuracy values of 77%, Precision 76% and Recall 72%, after the data balancing process obtained Accuracy values of 97%, Precision 83% and Recall 94%.

Keywords: *Sentiment analysis, Twitter, SVM, Oversampling, One Against Rest.*

DAFTAR ISI

HALAMAN PERSETUJUAN.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
HALAMAN PERNYATAAN	Error! Bookmark not defined.
KATA PENGANTAR	v
MOTTO	vi
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 <i>Sentiment analysis</i>	5
2.2 Twitter	5
2.3 <i>Text mining</i>	7
2.4 Klasifikasi.....	7
2.5 <i>Google Colaboratory</i>	7
2.6 Crawling	8
2.7 Python.....	8
2.8 Sastrawi	10

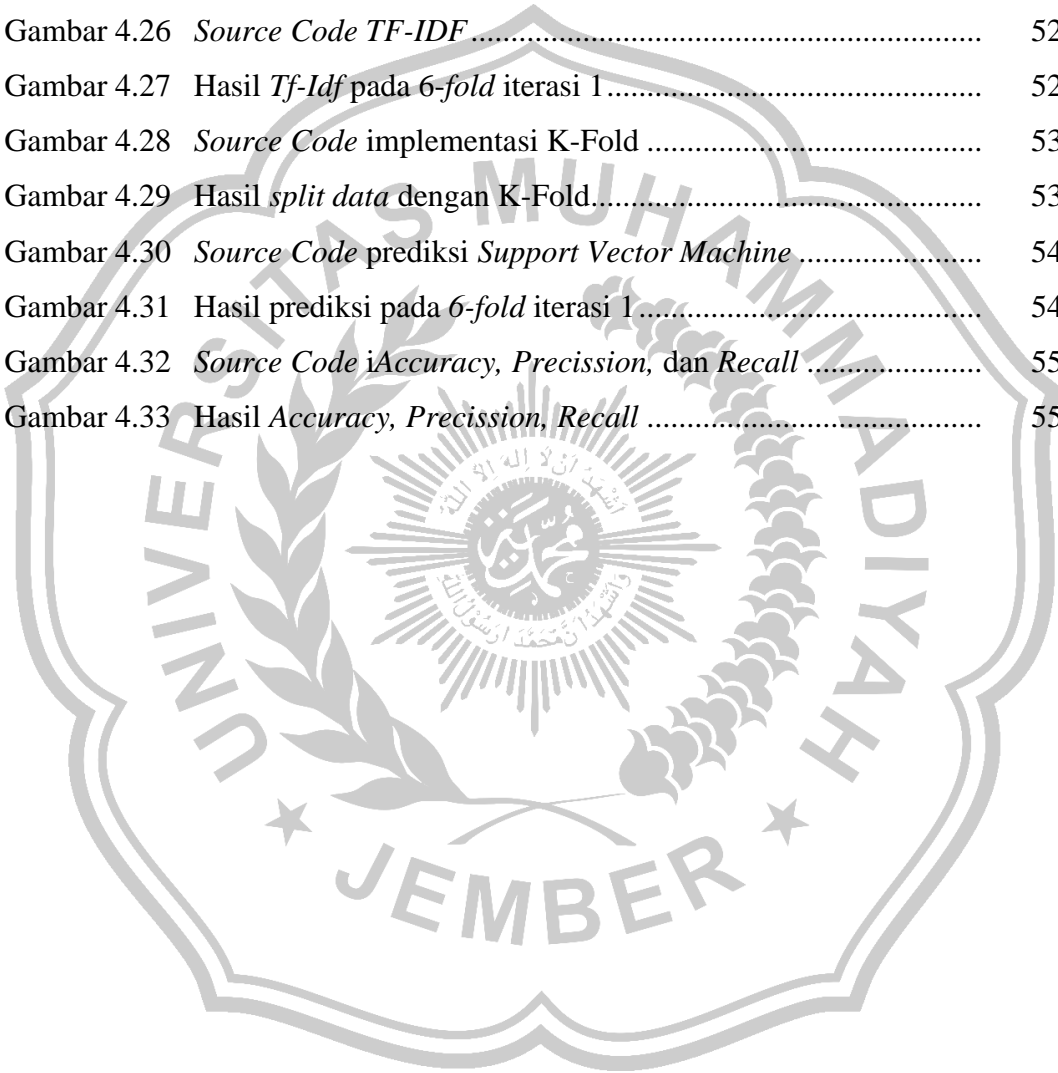
2.9	<i>Natural Language Toolkit (NLTK)</i>	10
2.10	<i>Text Preprocessing</i>	10
2.10.1	<i>Cleansing</i>	10
2.10.2	<i>Case Folding</i>	11
2.10.3	<i>Normalisasi Slangword</i>	12
2.10.4	<i>Tokenizing</i>	12
2.10.5	<i>Stopword Removal</i>	12
2.10.6	<i>Stemming</i>	13
2.11	<i>Resampling Data</i>	14
2.12	<i>Term Frequency-Inverse Document Frequency</i>	14
2.13	<i>Support Vector Machine</i>	16
2.14	<i>Overfitting dan Underfitting</i>	19
2.15	<i>K-fold Cross Validation</i>	20
2.16	<i>Confusion matrix</i>	20
2.17	<i>Penelitian terkait Support Vector Machine</i>	22
BAB III	METODOLOGI PENELITIAN	23
3.1	<i>Alur penelitian</i>	23
3.2	<i>Studi literatur</i>	23
3.3	<i>Observasi</i>	24
3.4	<i>Pengumpulan data</i>	24
3.5	<i>Labeling</i>	25
3.6	<i>Text preprocessing</i>	26
3.6.1	<i>Normalisasi Slangword</i>	27
3.6.2	<i>Stopwords removal</i>	27
3.6.3	<i>Stemming</i>	28

3.7	Resampling data	29
3.8	<i>Term Frequency-Invers Document Frequency</i>	29
3.9	<i>Support Vector Machine (SVM)</i>	33
3.10	<i>Confusion matrix</i>	39
BAB IV HASIL DAN PEMBAHASAN		40
4.1	Implementasi	40
4.1.1	Crawling Data	40
4.1.2	Import data	40
4.1.3	<i>Pre-Processing</i>	41
4.1.1.1	<i>Cleansing</i>	41
4.1.1.2	<i>Case Folding</i>	43
4.1.1.3	<i>Normalisasi Slangword</i>	44
4.1.1.4	<i>Tokenizing</i>	46
4.1.1.5	<i>Stopword Removal</i>	47
4.1.1.6	<i>Stemming</i>	49
4.1.4	<i>Oversampling</i>	50
4.1.5	Pembobotan <i>TF-IDF</i>	52
4.1.6	Implementasi <i>Support Vector Machine</i>	53
4.1.7	Uji Model	55
4.1.8	Evaluasi Model.....	59
BAB V KESIMPULAN DAN SARAN		64
5.1	Kesimpulan.....	64
5.2	Saran	64
DAFTAR PUSTAKA		65
LAMPIRAN.....		69

DAFTAR GAMBAR

Gambar 2.1	Overfitting dan Underfitting.....	19
Gambar 3.1	Diagram Alir Mendapatkan Twiter API.....	24
Gambar 3.2	Source Code Crawling Data	25
Gambar 3.3	Diagram Alir Text Pre-Processing	26
Gambar 3.4	Diagram alir normalisasi slangword.....	27
Gambar 3.5	Diagram Alir Proses Stopword Removal	28
Gambar 3.6	Diagram Alir Proses Stemming.....	28
Gambar 3.7	Diagram alir Tf-Idf.....	30
Gambar 3.8	Diagram alir perhitungan SVM.....	33
Gambar 4. 1	Hasil <i>Crawling</i> Data.....	40
Gambar 4.2	<i>Source Code Import data</i>	40
Gambar 4.3	Dataset.....	41
Gambar 4.4	<i>Source Code Cleansing</i>	41
Gambar 4.5	Data sebelum <i>Cleansing</i>	42
Gambar 4.6	Data sesudah <i>Cleansing</i>	42
Gambar 4.7	<i>Source Code Case Folding</i>	43
Gambar 4.8	Data sebelum <i>Case folding</i>	43
Gambar 4.9	Data sesudah <i>Case folding</i>	43
Gambar 4.10	<i>Source Code Normalisasi Slangword</i>	44
Gambar 4.11	Kamus <i>Slang Colloquial Indonesian Lexicon</i>	45
Gambar 4.12	Data sebelum <i>normalisasi slangword</i>	45
Gambar 4.13	Data sesudah <i>normalisasi slangword</i>	45
Gambar 4.14	<i>Source Code Tokenizing</i>	46
Gambar 4.15	Data sebelum <i>tokenizing</i>	46
Gambar 4.16	Data sesudah <i>tokenizing</i>	47
Gambar 4.17	<i>Source Code Stopword Removal</i>	47
Gambar 4.18	Kamus <i>Stopword</i>	48
Gambar 4.19	Data sebelum <i>Stopword Removal</i>	48

Gambar 4.20	Data setelah <i>Stopword Removal</i>	48
Gambar 4.21	<i>Source Code Stemming</i>	49
Gambar 4.22	Data sebelum <i>Stemming</i>	50
Gambar 4.23	Data sesudah <i>Stemming</i>	50
Gambar 4.24	<i>Source Code Oversampling</i>	50
Gambar 4.25	Data <i>Oversampling</i>	51
Gambar 4.26	<i>Source Code TF-IDF</i>	52
Gambar 4.27	Hasil <i>Tf-Idf</i> pada <i>6-fold</i> iterasi 1	52
Gambar 4.28	<i>Source Code</i> implementasi <i>K-Fold</i>	53
Gambar 4.29	Hasil <i>split data</i> dengan <i>K-Fold</i>	53
Gambar 4.30	<i>Source Code</i> prediksi <i>Support Vector Machine</i>	54
Gambar 4.31	Hasil prediksi pada <i>6-fold</i> iterasi 1	54
Gambar 4.32	<i>Source Code</i> <i>iAccuracy, Precision, dan Recall</i>	55
Gambar 4.33	Hasil <i>Accuracy, Precision, Recall</i>	55



DAFTAR TABEL

Tabel 3.1	Pelabelan data	26
Tabel 3.2	Nilai <i>Term frequency</i>	31
Tabel 3.3	Nilai <i>tf norm</i>	32
Tabel 3.4	Nilai dari <i>Df, Idf, dan Tf-idf</i>	33
Tabel 3.5	data Nilai Tf-Idf	34
Tabel 3.6	Nilai kernel Polynominal	35
Tabel 3.7	Nilai Matriks Hessian	35
Tabel 3.8	Nilai <i>Error rate</i> sampai iterasi ketiga	36
Tabel 3.9	Nilai dari delta alpha hingga iterasi ketiga	36
Tabel 3.10	Nilai alpha baru	37
Tabel 3.11	Nilai bias	37
Tabel 3.12	Perhitungan Hasil data uji	38
Tabel 3.13	<i>Confusion matrix</i>	39
Tabel 4.1	Hasil 2-fold <i>Cross Validation</i>	55
Tabel 4.2	Hasil 2-Fold iterasi 1	56
Tabel 4.3	Hasil 2-Fold iterasi 2	56
Tabel 4.4	Hasil 3-fold <i>Cross Validation</i>	56
Tabel 4.5	Hasil 3-Fold iterasi 1	56
Tabel 4.6	Hasil 3-Fold iterasi 2	56
Tabel 4.7	Hasil 3-Fold iterasi 3	57
Tabel 4.8	Hasil 4-fold <i>Cross Validation</i>	57
Tabel 4.9	Hasil 4-Fold iterasi 1	57
Tabel 4.10	Hasil 4-Fold iterasi 2	57
Tabel 4.11	Hasil 4-Fold iterasi 3	57
Tabel 4.12	Hasil 4-Fold iterasi 4	58
Tabel 4.13	Hasil 6-fold <i>Cross Validation</i>	58
Tabel 4.14	Hasil 6-Fold iterasi 1	58
Tabel 4.15	Hasil 6-Fold iterasi 2	58
Tabel 4.16	Hasil 6-Fold iterasi 3	59

Tabel 4.17	Hasil 6-Fold iterasi 4	59
Tabel 4.18	Hasil 6-Fold iterasi 5	59
Tabel 4.19	Hasil 6-Fold iterasi 6	59
Tabel 4.20	Accuracy data sebelum dan sesudah <i>Oversampling</i>	60
Tabel 4.21	Perhitungan Accuracy 6-fold iterasi ke-3	60
Tabel 4.22	<i>Precision 2-fold</i>	61
Tabel 4.23	<i>Precision 3-fold</i>	61
Tabel 4.24	<i>Precision 4-fold</i>	61
Tabel 4.25	<i>Precision 6-fold</i>	62
Tabel 4.26	<i>Recall 2-fold</i>	62
Tabel 4.27	<i>Recall 3-fold</i>	62
Tabel 4.28	<i>Recall 4-fold</i>	63
Tabel 4.29	<i>Recall 6-fold</i>	63
Tabel 4.30	Nilai <i>Recall</i> terbaik	63

