



**ANALISIS SENTIMEN PENGGUNA APLIKASI TWITTER
TERHADAP POLUSI UDARA DI JAKARTA DENGAN
MENGUNAKAN METODE
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

RAKAN YUVI ISPRADITYO

2010512104

**PROGRAM STUDI S1 SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
2024**



**ANALISIS SENTIMEN PENGGUNA APLIKASI TWITTER
TERHADAP POLUSI UDARA DI JAKARTA DENGAN
MENGUNAKAN METODE
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Mendapatkan Gelar Sarjana
Komputer**

RAKAN YUVI ISPRADITYO

2010512104

**PROGRAM STUDI S1 SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
2024**

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri.dan semua sumber yang dikutip maupun yang di rujuk telah saya nyatakan benar.

Nama : Rakan yuvi Ispradityo

NIM : 2010512104

Tanggal : 16 Januari 2025

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan di proses dengan ketentuan berlaku.

Jakarta ,16 Januari 2025

Yang menvatakan.



Rakan Yuvi ispradityo

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai Civitas akademik Universitas Pembangunan Nasional "Veteran" Jakarta, saya yang bertanda tangan di bawah ini.

Nama : Rakan Yuvi Ispradityo
NIM : 2010512104
Fakultas : Ilmu Komputer
Program Studi : SI - Sistem Informasi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional "Veteran" Jakarta Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalty Free Right) atas karya ilmiah saya yang berjudul:

ANALISIS SENTIMEN PENGGUNA APLIKASI TWITTER TERHADAP POLUSI UDARA DI JAKARTA DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional "Veteran" Jakarta berhak menyimpan, Mengalih data/formatkan, mengelola dalam bentuk pangkalan data (database), merawat dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Dibuat di : Jakarta
Pada Tanggal : 16 Januari 2024
Yang Menyatakan,



Rakan Yuvi Ispradityo

LEMBAR PERSETUJUAN

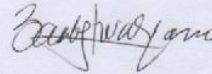
Nama : Rakan Yuvi Ispradityo
NIM : 2010512104
Program Studi : S1 Sistem Informasi
Judul : ANALISIS SENTIMEN PENGGUNA APLIKASI TWITTER TERHADAP POLUSI UDARA DI JAKARTA DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

Sebagai bagian persyaratan yang diperlukan untuk mengikuti Sidang Tugas Akhir/Skripsi pada Program Studi Sistem Informasi Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jakarta.

Menyetujui,

Dosen Pembimbing 1,

Dosen Pembimbing 2,

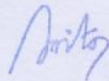


Ruth Mariana Bunga Wadu, S.Kom., MMSI

Bambang Tri Wahyono, S.Kom., M.Si.

Mengetahui,

Ketua Program Studi,



Anita Muliawati, S.Kom., M.T.I.

LEMBAR PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Rakan Yuvi Ispradityo

NIM : 2010512104

Program Studi : S1-Sistem Informasi

Judul Tugas Akhir : ANALISIS SENTIMEN PENGGUNA APLIKASI TWITTER TERHADAP POLUSI UDARA DI JAKARTA DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

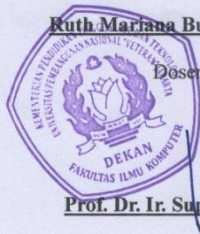
Telah berhasil dipertahankan dihadapan Tim penguji dan diterima sebagai bagian dari persyaratan yang diperlukan untuk memperoleh gelar Sarjana Ilmu Komputer pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jakarta.

Ati Zaidiah, S.Kom, MTL.
Dosen Penguji 1

Nindy Irzayika, S.SI., M.T.
Dosen Penguji 2

Ruth Mariana Bunga Wadu S.Kom., MMSI
Dosen Pembimbing 1

Bambang Tri Wahyono, S.Kom., M.Si.
Dosen Pembimbing 2



Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM
Dekan

Anita Muliawati, S.Kom., MTL.
Koordinator Program Studi S1 Sistem Informasi

Ditetapkan di : Jakarta
Tanggal Ujian : 16 Desember 2024

**ANALISIS SENTIMEN PENGGUNA APLIKASI TWITTER
TERHADAP POLUSI UDARA DI JAKARTA DENGAN MENGGUNAKAN
METODE
SUPPORT VECTOR MACHINE (SVM)**

Rakan Yuvi Ispradityo

Abstrak

Pada era teknologi yang semakin maju, media sosial, terutama Twitter, telah menjadi salah satu platform utama untuk menyuarakan opini dan isu-isu sosial. Salah satu isu yang kerap dibicarakan di Twitter adalah polusi udara di Jakarta, yang dianggap sebagai salah satu masalah lingkungan terbesar di kota tersebut. Penelitian ini bertujuan untuk menganalisis sentimen pengguna Twitter terhadap polusi udara di Jakarta menggunakan metode Support Vector Machine (SVM). Data yang digunakan dalam penelitian ini berasal dari 2022 tweet berbahasa Indonesia yang dikumpulkan dari September 2023 hingga Agustus 2024. Proses pre-processing dilakukan untuk membersihkan dan menyiapkan data, kemudian dilakukan pelabelan sentimen menjadi positif atau negatif. Metode SVM dengan kernel linear digunakan untuk memprediksi sentimen, dengan pembagian data latih dan uji secara acak. Hasil penelitian menunjukkan bahwa sentimen negatif mendominasi, dengan 1884 tweet bersentimen negatif dan 111 tweet bersentimen positif. Penelitian ini menegaskan bahwa SVM memiliki performa yang baik dalam klasifikasi sentimen, dan polusi udara di Jakarta menjadi topik yang mendapatkan perhatian signifikan dari masyarakat.

Kata Kunci : Twitter, polusi udara, Jakarta, analisis sentimen, Support Vector Machine (SVM), media sosial, data pre-processing, klasifikasi sentimen.

**ANALISIS SENTIMEN PENGGUNA APLIKASI TWITTER
TERHADAP POLUSI UDARA DI JAKARTA DENGAN MENGGUNAKAN
METODE
SUPPORT VECTOR MACHINE (SVM)**

RAKAN YUVI ISPRADITYO

Abstract

In this era of advanced technology, social media, especially Twitter, has become one of the main platforms for voicing opinions and social issues. One issue that is often discussed on Twitter is air pollution in Jakarta, which is considered one of the biggest environmental problems in the city. This study aims to analyze the sentiment of Twitter users towards air pollution in Jakarta using the Support Vector Machine (SVM) method. The data used in this study comes from 2022 Indonesian tweets collected from September 2023 to August 2024. Pre-processing is done to clean and prepare the data, then sentiment labeling is done to be positive or negative. SVM method with linear kernel was used to predict sentiment, with random division of training and test data. The results showed that negative sentiment dominated, with 1884 tweets with negative sentiment and 111 tweets with positive sentiment. This study confirms that SVM performs well in sentiment classification, and air pollution in Jakarta is a topic that has received significant attention from the public.

Keywords : Twitter, air pollution, Jakarta, sentiment analysis, Support Vector Machine (SVM), social media, data pre-processing, sentiment classification.

KATA PENGANTAR

Puji syukur penulis panjatkan atas kehadiran Allah SWT, yang telah memberikan rahmat, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan penulisan Proposal Tugas Akhir ini dengan baik. Penyusunan Proposal Tugas Akhir ini dilakukan dengan tujuan memenuhi persyaratan yang ditetapkan untuk penyusunan Tugas Akhir/Skripsi.

Dalam proses penyelesaian proposal ini merupakan hasil dari perjuangan, kerja keras, dan bimbingan dari berbagai pihak. Demikian, segala rasa hormat dan terima kasih disampaikan setinggi – tingginya kepada :

1. Kepada Allah SWT atas rahmat, hidayah serta karunia-Nya.
2. Kedua Orang Tua yang telah memberikan dukungan sepenuh hati.
3. Bapak Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM, selaku Dekan Fakultas Ilmu Komputer UPNVJ.
4. Ibu Anita Muliawati, S.Kom., MTI., selaku Ketua Jurusan S1 Sistem Informasi.
5. Ibu Ruth Mariana Bunga Wadu S.Kom., MMSI, yang telah memberikan arahan, bimbingan, serta dukungan yang luar biasa selama proses penulisan skripsi ini.
6. Pak Bambang Tri Wahyono, S.Kom., M.SI., selaku Dosen Pembimbing 2.
7. Seluruh pihak yang tidak dapat disebutkan satu – persatu.

Penulis menyadari bahwa skripsi ini jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk perbaikan di masa mendatang. Akhir kata, semoga skripsi ini dapat bermanfaat bagi pembaca yang ingin mengetahui lebih dalam tentang topik yang dibahas. penulis berharap hasil penelitian ini dapat menjadi sumbangsih kecil dalam pengembangan ilmu pengetahuan di bidang yang bersangkutan.

Jakarta, 1 Mei 2024
Rakan Yuvi Ispradityo

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iv
LEMBAR PERSETUJUAN.....	v
ABSTRAK.....	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN.....	xiv
DAFTAR SIMBOL.....	xv
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 Luaran yang diharapkan	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Analisis Sentimen	6
2.2 Twitter.....	6
2.3 Twitter Crawling.....	6
2.4 Polusi Udara.....	7
2.5 <i>Data Mining</i>	7
2.6 <i>Text Mining</i>	7
2.7 <i>Pre-Processing Data</i>	8
2.7.1 <i>Data Cleaning</i>	8
2.7.2 <i>Case Folding</i>	9
2.7.3 <i>Normalization</i>	9
2.7.4 <i>Stemming</i>	10
2.7.5 <i>Stopword</i>	11
2.8 Pelabelan Lexicon Based.....	11
2.9 Pembobotan Kata <i>Term Frequency – Inverse Document Frequency</i> (TF – IDF).....	12
2.10 <i>Algoritma Support Vector Machine</i> (SVM)	13
2.11 Evaluasi.....	15
2.12 Python	16
2.13 Penelitian Terdahulu.....	16

BAB III	METODOLOGI PENELITIAN.....	20
3.1	Alur penelitian	20
3.1.1	Identifikasi Masalah.....	21
3.1.2	Studi Literatur	21
3.1.3	Pengumpulan Data.....	21
3.1.4	Pelabelan Data	21
3.1.5	<i>Pre-Processing</i> Data.....	22
3.1.6	Pelabelan Lexicon Based.....	23
3.1.7	Pembobotan Kata TF-IDF	23
3.1.8	Klasifikasi	24
3.1.9	Evaluasi.....	24
3.1.10	Hasil Klasifikasi.....	24
3.2	Alat Bantu Penelitian.....	25
3.3	Jadwal Kegiatan.....	26
BAB IV	HASIL DAN PEMBAHASAN.....	27
4.1	Pengumpulan Data.....	27
4.2	<i>Preprocessing</i> Data.....	28
4.2.1	<i>Data Cleaning</i>	30
4.2.2	<i>Case Folding</i>	33
4.2.3	Normalization	35
4.2.4	Stemming.....	37
4.2.5	Stopword.....	39
4.3	<i>Labelling</i>	41
4.4	Pembobotan Kata.....	43
4.5	Split Data	45
4.6	Klasifikasi dengan Support Vector Machine (SVM).....	47
4.7	Hasil Evaluasi Model.....	48
4.8	Labelling dengan SVM.....	50
4.9	Hasil dan Visualisasi.....	53
4.10	Sentimen Negatif	53
4.11	Sentimen Positif.....	54
4.12	Visualisasi Word Cloud Sentimen Negatif.....	55
4.13	Implementasi Pengembangan Sistem	58
4.13.1	Tampilan Halaman Landing page.....	59
4.13.2	Tampilan Halaman Dashboard	60
BAB V	KESIMPULAN	62
5.1	Kesimpulan.....	62
5.2	Saran	62
DAFTAR PUSTAKA		67
RIWAYAT HIDUP		
LAMPIRAN		

DAFTAR GAMBAR

Gambar 2 1 Hyperlane memisahkan dua kelas	14
Gambar 3 1 Alur Penelitian	20
Gambar 4 1 Sintaks Crawling Data	27
Gambar 4 2 Sintaks Filter Data Relevan	27
Gambar 4 3 Hasil Crawling Data	28
Gambar 4 4 Sintaks Sastrawi dan Emoji	29
Gambar 4 5 Sintaks NLP dan NLTK.....	29
Gambar 4 6 Sintaks Data Cleaning.....	32
Gambar 4 7 Sintaks Case Folding	33
Gambar 4 8 Sintaks Normalization	35
Gambar 4 9 Sintaks Stemming	37
Gambar 4 10 Sintaks Stopwording.....	39
Gambar 4 11 Kamus Inset lexicon	42
Gambar 4 12 Sintaks Split Data	45
Gambar 4 13 Hasil Confusion Matrix	49
Gambar 4 14 Sintaks Pelabelan SVM	51
Gambar 4 15 Persentase Hasil Pelabelan	52
Gambar 4 16 Top 10 Kata Negatif	53
Gambar 4 17 Top 10 Kata Positif.....	54
Gambar 4 18 Visualisasi Wordcloud Negatif.....	55
Gambar 4 19 Visualisasi Wordcloud Positif	57
Gambar 4 20 Halaman Landing Page.....	59
Gambar 4 21 Halaman Dashboard Bagian 1	60
Gambar 4 22 Halaman Dashboard Bagian 2	60

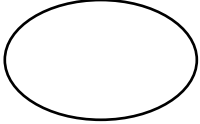


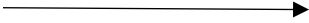
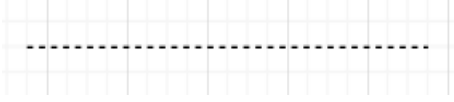
DAFTAR TABEL

Tabel 2 1 Contoh Casefolding	9
Tabel 2 2 Contoh Normalization	10
Tabel 2 3 Contoh Stemming	10
Tabel 2 4 Contoh Stopword	11
Tabel 2 5 Persamaan Setiap Kernel SVM	15
Tabel 2 6 Confusion Matrix	15
Tabel 2 7 Penelitian Terdahulu	16
Tabel 3 1 Jadwal Penelitian	26
Tabel 4 1 Penerapan Data Cleaning	32
Tabel 4 2 Penerapan Case Folding	34
Tabel 4 3 Penerapan Normalization	36
Tabel 4 4 Hasil Penerapan Stemming	38
Tabel 4 5 Hasil Penerapan Stopword	40
Tabel 4 6 Hasil Pelabelan Kata	42
Tabel 4 7 Hasil Pembobotan Kata	44
Tabel 4 8 Banyak Data Latih dan Data Uji	45
Tabel 4 9 Perbandingan Split Data	46
Tabel 4 10 Kata Koefisien Positif	47
Tabel 4 11 Kata Koefisien negatif	47
Tabel 4 12 Hasil Evaluasi Model	50
Tabel 4 13 Hasil Implementasi Labelling SVM	51

DAFTAR LAMPIRAN

Lampiran 1 Import Library.....	69
Lampiran 2 Crawling Data	69
Lampiran 3 Case folding	70
Lampiran 4 Cleaning Data.....	72
Lampiran 5 Normalization.....	74
Lampiran 6 Stemming	76
Lampiran 7 Stopword	78
Lampiran 8 Inset Lexicon Based	80
Lampiran 9 Pembobotan dengan TF-IDF.....	83
Lampiran 10 Korpus katabaku untuk proses normalisasi.....	85

DAFTAR SIMBOL

No	Gambar	Nama	Keterangan
1		Terminator Start/End	Tanda awalan atau akhir dari alur suatu flowchart.
2		Process	Menggambarkan pengolahan yang dilakukan oleh Komputer.
3		Input/Output	Merupakan gambaran dari input/output proses.
4		Flow Line	Merupakan gambaran aliran suatu data.
5		Dependency	Menghubungkan elemen Indeodent dengan elemn non- Independent