

**ANALISIS SENTIMEN PILKADA JATIM 2024 MENGGUNAKAN
METODE RANDOM FOREST PADA WEBSITE BERITA DAN
ULASAN PUBLIK DI MEDIA SOSIAL TWITTER**

PRAKTIK KERJA LAPANGAN



Oleh :

ARSA CAHAYA PRADIPTA	NPM : 22082010015
NABILA AHLISYA FATIR	NPM : 22082010017
AL-FAIZ AZZAM ARYAPUTRA	NPM : 22082010022

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
Universitas Pembangunan Nasional “Veteran” Jawa Timur
S U R A B A Y A
2024**

HALAMAN PENGESAHAN

Judul : ANALISIS SENTIMEN PILKADA JATIM 2024
MENGUNAKAN METODE RANDOM FOREST PADA
WEBSITE BERITA DAN ULASAN PUBLIK DI MEDIA SOSIAL
TWITTER

Oleh : ARSA CAHAYA PRADIPTA NPM: 22082010015
NABILA AHLISYA FATIR NPM: 22082010017
AL-FAIZ AZZAM ARYAPUTRA NPM: 22082010022

Menyetujui,

Pembimbing



Anita Wulansari, S.Kom, M.Kom

Pembimbing Lapangan



M. Dheo Fuady

NPT/NIP. 19871015 202203 2005

Mengetahui,

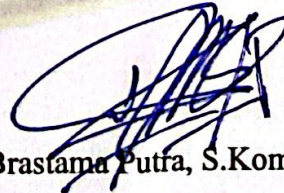
Dekan
Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT

NPT/NIP. 19681126 199403 2 001

Koordinator Program Studi
Sistem Informasi



Agung Brastama Putra, S.Kom, M.Kom

NPT/NIP. 19851124 2021211 003

Isi
dikorupsi May 71

ABSTRAK

Praktik Kerja Lapangan (PKL) ini bertujuan untuk menganalisis sentimen publik terhadap Pilkada Jatim 2024 menggunakan metode *Random Forest* pada data yang dikumpulkan dari berita daring dan media sosial Twitter. Penelitian ini dilatarbelakangi oleh pentingnya pemahaman mendalam terhadap opini publik dalam konteks politik, khususnya menjelang Pilkada, untuk mendukung pengambilan keputusan berbasis data. Dengan menggunakan pendekatan analisis sentimen, penelitian ini mengelompokkan opini publik ke dalam tiga kategori sentimen utama, yaitu positif, netral, dan negatif, untuk memberikan wawasan yang berguna mengenai persepsi masyarakat terhadap kandidat dan isu-isu utama dalam Pilkada Jatim 2024.

Penelitian ini melibatkan beberapa tahapan yang dimulai dengan perencanaan, pengumpulan data melalui teknik scraping, pra-pemrosesan data, pemodelan dan analisis sentimen, hingga evaluasi model dan visualisasi hasil. Data dikumpulkan dari dua sumber utama: berita daring yang diambil dari portal berita terpercaya dan media sosial Twitter menggunakan API untuk memperoleh tweet yang relevan dengan Pilkada Jatim 2024. Selama proses pra-pemrosesan, data dibersihkan dari elemen-elemen yang tidak relevan, seperti simbol, tanda baca, URL, dan emotikon. Selain itu, dilakukan tokenisasi dan stemming untuk menyederhanakan analisis data.

Data awal yang dikumpulkan menunjukkan ketidakseimbangan distribusi sentimen, dengan sentimen netral mendominasi. Untuk mengatasi masalah ini, metode *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan untuk menyeimbangkan distribusi data, dengan menambah data sintesis pada kategori sentimen positif dan negatif. Hasilnya, model dapat dilatih secara optimal untuk memprediksi semua kategori sentimen dengan lebih akurat.

Metode *Random Forest* dipilih sebagai algoritma utama karena kemampuannya dalam menangani data tidak terstruktur, seperti teks ulasan publik dan berita daring, serta tingkat akurasi yang tinggi dalam klasifikasi. Model ini berhasil mencapai performa evaluasi yang sempurna, dengan akurasi, *precision*, *recall*, dan *F1-score* sebesar 1.00 untuk semua kategori sentimen. Evaluasi yang dilakukan menunjukkan bahwa metode ini sangat efektif dalam mengklasifikasikan sentimen publik tanpa kesalahan.

Hasil analisis menunjukkan dominasi sentimen netral dalam pemberitaan Pilkada Jatim 2024, dengan peningkatan signifikan pada bulan-bulan mendekati Pilkada. Analisis lebih lanjut terhadap kandidat menunjukkan bahwa Paslon 2 mendominasi pemberitaan di semua kategori sentimen, baik positif, netral, maupun negatif, yang mencerminkan perhatian media dan publik yang lebih besar terhadap pasangan tersebut.

Visualisasi data dilakukan untuk memberikan wawasan yang mudah dipahami oleh pemangku kepentingan. Grafik dan tabel yang dihasilkan memaparkan tren sentimen secara umum dan distribusi sentimen berdasarkan pasangan calon. Wawasan yang diperoleh dari analisis ini dapat dimanfaatkan oleh pemangku kepentingan politik untuk menyusun strategi kampanye yang lebih efektif dan responsif terhadap isu strategis. Selain itu, hasil analisis juga memberikan gambaran mengenai isu-isu utama yang menjadi perhatian masyarakat, sehingga dapat membantu para kandidat dalam menyusun pesan-pesan kampanye yang lebih relevan dan sesuai dengan ekspektasi publik.

Penelitian ini tidak hanya memberikan kontribusi akademik dengan mengaplikasikan teori dan teknologi analisis data dalam konteks nyata, tetapi juga menghasilkan wawasan praktis yang relevan untuk pengambilan keputusan berbasis data. Proses pengumpulan data yang sistematis, didukung oleh teknik pra-pemrosesan yang cermat, memastikan bahwa hasil analisis dapat diandalkan dan bermakna.

Kedepannya, beberapa saran untuk penelitian lanjutan meliputi penggunaan metode analisis yang lebih canggih, seperti algoritma *Deep Learning*, untuk meningkatkan akurasi dan efisiensi dalam pemrosesan data. Selain itu, perluasan sumber data dengan mengintegrasikan platform media sosial lain, seperti *Facebook* atau *Instagram*, dapat memberikan representasi opini publik yang lebih komprehensif. Hasil analisis sentimen juga dapat diintegrasikan langsung ke dalam strategi kampanye atau kebijakan publik untuk merespons isu strategis secara lebih efektif.

Secara keseluruhan, kegiatan PKL ini membuktikan bahwa metode *Random Forest* mampu mengklasifikasikan sentimen publik dengan akurasi yang tinggi dan menghasilkan wawasan yang berguna. Penelitian ini memberikan solusi berbasis teknologi yang sistematis untuk memahami opini publik dalam konteks politik, yang pada akhirnya dapat mendukung pengambilan keputusan yang lebih baik oleh berbagai pemangku kepentingan. Selain itu, analisis ini juga memberikan nilai tambah bagi mahasiswa dalam hal peningkatan kompetensi di bidang teknologi dan analisis data, yang relevan dengan tantangan industri di era digital.

Kegiatan PKL ini tidak hanya berperan sebagai sarana pembelajaran, tetapi juga menunjukkan bahwa data dan teknologi dapat dioptimalkan untuk memahami dinamika sosial-politik dengan lebih baik. Ini menjadi bukti nyata bahwa integrasi teknologi, data, dan analisis dapat memberikan dampak positif bagi pengambilan keputusan di berbagai sektor.

Kata Kunci: *Analisis Sentimen, Random Forest, Pilkada Jatim 2024, Media Sosial, Scraping Data.*

KATA PENGANTAR

Puji syukur kami panjatkan kehadirat Allah SWT atas rahmat dan hidayah-Nya kami dapat menyelesaikan laporan Praktik Kerja Lapangan (PKL) yang berjudul “Analisis Sentimen Pilkada Jatim 2024 Menggunakan Metode Random Forest pada Website Berita dan Ulasan Publik di Media Sosial Twitter” dengan tepat waktu. Adapun maksud dan tujuan kami dalam menyelesaikan laporan PKL ini untuk memenuhi sebagian persyaratan guna menyelesaikan program studi independen bersertifikat di perusahaan Masyarakat Industri Kreatif Teknologi Informasi dan Komunikasi Indonesia (MIKTI). Kami menyadari dalam penyelesaian laporan ini banyak pihak yang telah membantu, oleh karena itu pada kesempatan ini kami mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Masyarakat Industri Kreatif Teknologi Informasi dan Komunikasi Indonesia (MIKTI) selaku perusahaan tempat kami menjalani program Studi Independen Bersertifikat, atas kesempatan, dukungan, dan pengalaman berharga yang diberikan selama program berlangsung.
2. M. Dheo Fuady selaku Mentor DA-2 MIKTI, atas dukungan dan bimbingannya selama pelaksanaan program.
3. Ibu Anita Wulansari, S.Kom, M.Kom. selaku dosen pembimbing PKL yang selalu memberikan saran dalam penyusunan laporan ini.
4. Teman-teman, serta pihak - pihak yang tidak dapat kami sebutkan satu persatu yang telah mendukung dan membantu dalam menyelesaikan laporan PKL.

Semoga laporan yang telah kami susun ini memberikan manfaat baik bagi penyusun, pemangku kebijakan, pembaca, dan masyarakat luas nantinya. Kami menyadari bahwa laporan ini masih jauh dari kata sempurna. Oleh karena itu, kami sangat mengharapkan kritik dan saran yang bersifat membangun dari berbagai pihak guna meningkatkan kualitas laporan kami di masa mendatang.

Surabaya, 6 Januari 2025

Penyusun

DAFTAR ISI

HALAMAN PENGESAHAN	i
ABSTRAK	ii
KATA PENGANTAR.....	iv
DAFTAR ISI	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN	x
BAB I	1
PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan PKL	2
1.4 Manfaat	2
BAB II.....	4
GAMBARAN UMUM TEMPAT PKL	4
2.1 Profil Organisasi	4
2.2 Visi dan Value Organisasi	4
2.2.1 Visi Organisasi.....	4
2.2.2 Value Organisasi	4
2.3 Struktur Organisasi	5
2.4 Bidang Usaha Organisasi	6
BAB III.....	7
PELAKSANAAN PKL.....	7
3.1 Tinjauan Pustaka	7
3.1.1 Analisis Sentimen	7
3.1.2 Pilkada dan Media Sosial.....	7
3.1.3 Python	8
3.1.4 Visual Studio Code	9

3.1.5 Google Colaboratory.....	9
3.1.6 Data Crawling	10
3.1.7 Data Scraping.....	10
3.1.8 Metode Random Forest.....	11
3.2 Waktu dan Tempat Pelaksanaan PKL.....	11
3.2.1 Tempat dan Waktu Penelitian.....	11
3.2.2 Pelaksanaan PKL	12
BAB IV	16
HASIL DAN PEMBAHASAN.....	16
4.1 Analisis Kebutuhan	16
4.1.1 Identifikasi Masalah.....	16
4.1.2 Penentuan Metode Analisis.....	16
4.2 Hasil Kegiatan.....	19
4.2.1 Pengumpulan Data	19
4.2.2 Pra-Pemrosesan Data	24
4.2.3 Transformasi Teks	34
4.2.4 Data Latih dan Data Uji	36
4.2.5 Penyeimbangan Data	37
4.2.6 Penerapan Model <i>Random Forest</i>	38
4.2.7 Evaluasi Model	40
4.2.8 Interpretasi dan Visualisasi Hasil.....	41
BAB V.....	65
PENUTUP	65
5.1 Kesimpulan	65
5.2 Saran	66
DAFTAR PUSTAKA	67
LAMPIRAN	69

DAFTAR TABEL

Tabel 3.1 Daftar Pekerjaan yang Dilakukan	14
Tabel 4.1 <i>Problem Questions</i>	16

DAFTAR GAMBAR

Gambar 2.1 Struktur Kepengurusan Organisasi MIKTI	5
Gambar 3.1 <i>Flowchart</i> Tahapan Pelaksanaan	12
Gambar 4.1 Data Hasil Crawling Menggunakan API Twitter	19
Gambar 4.2 Data Hasil Crawling Menggunakan Tweet Harvest.....	20
Gambar 4.3 Data Hasil Scraping Pada Website Tribunnews.....	21
Gambar 4.4 Data Hasil Scraping pada Website Detikcom	22
Gambar 4.5 Dataset Pilkada Jatim 2024	22
Gambar 4.6 Dataset Label Map Pilkada Jatim 2024	23
Gambar 4.7 <i>Analysis Tools</i>	24
Gambar 4.8 <i>Gathering Data</i>	24
Gambar 4.9 Output Pengecekan <i>Missing Values</i>	25
Gambar 4.10 Output Konversi Kolom	26
Gambar 4.11 <i>Data Cleaning</i>	27
Gambar 4.12 Output <i>Pre Processing Text</i>	28
Gambar 4.13 <i>Labeling Data</i>	28
Gambar 4.14 <i>Library Implementation</i>	29
Gambar 4.15 Output <i>Tokenizing</i>	30
Gambar 4.16 Output <i>Stopwords</i>	32
Gambar 4.17 Output <i>Stemming</i>	34
Gambar 4.18 Output Transformasi Teks dan Analisis Sentimen.....	35
Gambar 4.19 Dimensi Data Latih dan Data Uji	36
Gambar 4.20 RandomForestClassifier	40
Gambar 4.21 Output Model Evaluation	41
Gambar 4.22 Diagram Kemunculan Berita untuk Setiap Paslon	42
Gambar 4.23 Frekuensi Kemunculan Berita untuk Setiap Paslon	44
Gambar 4.24 Tren Pemberitaan Pilkada Jatim 2024.....	46
Gambar 4.25 Tren Pemberitaan untuk Tiap Paslon per Bulan.....	49
Gambar 4.26 Kata yang Banyak Disebut dalam Berita Pilkada	51

Gambar 4.27 Kata yang Paling Banyak Disebut untuk Paslon 1	53
Gambar 4.28 Kata yang Paling Banyak Disebut untuk Paslon 2	55
Gambar 4.29 Kata yang Paling Banyak Disebut untuk Paslon 3	57
Gambar 4.30 Funnel-Chart Distribusi Sentimen	58
Gambar 4.31 Sentimen Berita per Bulan	60
Gambar 4.32 Grafik Sentimen Pemberitaan Tiap Bulan per Paslon	64

DAFTAR LAMPIRAN

Lampiran 1. Konsultasi Pertama Bersama Pembimbing Lapangan.....	69
Lampiran 2. Konsultasi Kedua Bersama Pembimbing Lapangan.....	69
Lampiran 3. Konsultasi Ketiga Bersama Pembimbing Lapangan	70
Lampiran 4. <i>Final Presentation</i>	70