

**Analisis Sentimen Menggunakan Metode *Random Forest*, *SVM*
dan *XGBoost* Pada Tayangan TV Berdasarkan Ulasan
Masyarakat di Media Sosial Twitter**

PRAKTEK KERJA LAPANGAN



Oleh:

1. **RIO ALGHANIY PUTRA** (22082010012)
2. **ADAM IDHOFI RAKASIWI** (22082010026)
3. **SALSABILA PUTRI AZZAHRA** (22082010079)

**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 MENGGUNAKAN METODE
RANDOM FOREST, SVM, DAN XGBOOST PADA
TAYANGAN TV BERDASARKAN ULASAN MASYARAKAT
DI MEDIA SOSIAL TWITTER

Oleh : RIO ALGHANIY PUTRA NPM 22082010012
ADAM IDHOFI RAKASIWI NPM 22082010026
SALSABILA PUTRI AZZAHRA NPM 22082010079

Menyetujui,

Dosen Pembimbing
PKL



Asif Faroqi, S.Kom, M.Kom
NPT/NIP 198705192018031001

Pembimbing Lapangan



Hadhlan Adha

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

ABSTRAK

Dalam era digital yang berkembang pesat, media sosial seperti Twitter menjadi platform utama bagi masyarakat dalam menyampaikan opini dan ulasan terhadap tayangan televisi. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap tayangan televisi berdasarkan ulasan yang diunggah di Twitter dengan menerapkan metode machine learning, yaitu Random Forest, Support Vector Machine (SVM), dan XGBoost. Ketiga metode ini dipilih karena kemampuan mereka dalam mengklasifikasikan data teks secara akurat. Proses analisis melibatkan beberapa tahapan utama, yaitu scraping data dari Twitter, data cleansing, labeling, tokenizing, stopword removal, serta penerapan teknik visualisasi data menggunakan word cloud. Model Random Forest, SVM, dan XGBoost diterapkan untuk mengklasifikasikan data sentimen yang telah diproses, dengan evaluasi performa menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa metode XGBoost memiliki performa yang unggul dibandingkan dengan Random Forest dan SVM dalam hal akurasi dan kecepatan proses pada dataset yang digunakan. Temuan ini memberikan wawasan yang bermanfaat bagi stasiun televisi dalam mengevaluasi opini penonton secara lebih efektif dan berbasis data, yang dapat mendukung peningkatan kualitas program serta pengambilan keputusan strategis yang lebih tepat sasaran.

Kata kunci: Analisis Sentimen, Twitter, *Random Forest*, *Support Vector Machine (SVM)*, *XGBoost*, Tayangan Televisi

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya, sehingga kami dapat menyelesaikan *Project Capstone* yang berjudul "*Analisis Sentimen Menggunakan Metode Random Forest, SVM, dan XGBoost pada Tayangan TV Berdasarkan Ulasan Masyarakat di Media Sosial Twitter*".

Kami ingin mengucapkan rasa terima kasih yang sebesar-besarnya kepada:

1. Orang tua kami, atas doa, dukungan, dan kasih sayang yang senantiasa menjadi sumber kekuatan dalam menyelesaikan *Project Capstone* ini.
2. Prof. Dr. Ir. Akhmad Fauzi, MMT., IPU., CHRA, selaku Rektor Universitas Pembangunan Nasional "Veteran" Jawa Timur, atas dukungannya terhadap program studi yang kami jalani.
3. Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur, atas motivasi dan arahnya yang sangat berarti.
4. Bapak/Ibu Dosen Universitas Pembangunan Nasional "Veteran" Jawa Timur, atas bimbingan dan ilmunya yang sangat bermanfaat selama proses pembelajaran.
5. Kak Fadhlán dan Kak Dheo, selaku mentor pembelajaran, yang dengan sabar dan jelas menyampaikan materi analisis data dan machine learning sehingga kami mampu memahami konsep dengan baik.

Kami berharap hasil dari *Project Capstone* ini dapat memberikan kontribusi positif bagi pengembangan analisis data dan pengolahan bahasa alami di dunia akademis maupun industri. Kritik dan saran yang membangun sangat kami harapkan demi penyempurnaan karya ini di masa mendatang.

Surabaya, 23 Desember 2024

Tim Penyusun

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	ii
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vii
DAFTAR LAMPIRAN	viii
BAB I	1
PENDAHULUAN	1
1.1 Latar belakang	1
1.3 Tujuan PKL	3
1.4 Manfaat.....	3
BAB II	5
GAMBARAN UMUM TEMPAT PKL	5
2.1 Gambaran Umum Organisasi	5
2.1.1 Visi Organisasi.....	5
2.1.2 Misi Organisasi.....	5
2.1.3 Fokus Utama.....	5
BAB III	8
PELAKSANAAN PKL	8
3.1 Tinjauan Pustaka	8
3.2 Waktu dan Tempat Pelaksanaan PKL.....	13
3.3 Tahapan Analisis Sentimen	13
3.3.1 Scraping data dan crawling data.....	13
3.3.2 Implementasi Library Pendukung	15
3.3.3 Load data	16
3.3.4 Data Cleansing.....	16
3.3.5 Labeling.....	18
3.3.6 Tokenizing	18
3.3.7 Stopword removal	19
3.3.8 Visualisasi Data.....	19
3.3.9 Wordcloud	20
3.3.10 Modeling dengan Random Forest, SVM, XGBoost	21
3.4 Pembagian Tugas Antar Kelompok	22
BAB IV	24
HASIL DAN PEMBAHASAN	24
4.1 Deskripsi Project	24
4.2 Hasil	25

4.3.1 Library Pendukung	25
4.3.2 Load Data	27
4.3.3 Scraping Data	28
4.3.4 Cleansing dan Pre-processing Data	30
4.3.5 Tokenizing	34
4.3.6 Lemmatization	34
4.3.7 Stemming.....	35
4.3.8 Labeling	35
4.3.9 Wordcloud	37
4.3.10 Distribusi data.....	40
4.3.11 Modeling.....	42
4.3.12 Fungsi fold.....	44
4.3.13 Model Random Forest	45
4.3.14 Model SVM.....	47
4.3.15 Model XGBoost	49
4.3.16 Prediksi.....	52
4.3.17 Visualisasi.....	55
BAB V	63
PENUTUP.....	63
5.1 Kesimpulan	63
5.2 Saran	64
DAFTAR PUSTAKA	66
LAMPIRAN.....	68

DAFTAR GAMBAR

Gambar 2.1 Struktur Organisasi MIKTI	6
Gambar 4.1 Install library	26
Gambar 4.2 Import library.....	26
Gambar 4.3 Download pustaka NLTK.....	27
Gambar 4.4 Load dataset.....	28
Gambar 4.5 Crawling Data.....	30
Gambar 4.6 Output fungsi df.info().....	30
Gambar 4.7 Output fungsi df.describe().....	31
Gambar 4.8 Output fungsi df.isnull().sum()	31
Gambar 4.9 Output duplikat data	31
Gambar 4.10 Cek nilai null	32
Gambar 4.11 Proses cleansing text karakter.....	32
Gambar 4.12 Output cleansing text karakter.....	32
Gambar 4.13 Proses cleansing text typo	33
Gambar 4.14 Proses stopwords	34
Gambar 4.15 Proses tokenizing.....	34
Gambar 4.16 Proses lemmatization	35
Gambar 4.17 Proses dan output stemming.....	35
Gambar 4.18 Proses labeling.....	36
Gambar 4.19 Output hasil labeling.....	37
Gambar 4.20 Proses wordcloud sentimen netral	38
Gambar 4.21 Proses wordcloud sentimen positif.....	38
Gambar 4.22 Proses wordcloud sentimen negatif.....	39
Gambar 4.23 Wordcloud sentimen netral.....	39
Gambar 4.24 Wordcloud sentimen positif.....	39
Gambar 4.25 Wordcloud sentimen negatif.....	40
Gambar 4.26 Distribusi label sentimen	40
Gambar 4.27 Proses visualisasi distribusi sentimen.....	40
Gambar 4.28 Visualisasi distribusi sentimen	41
Gambar 4.29 Proses visualisasi distribusi sentimen funnel chart	41
Gambar 4.30 Visualisasi distribusi sentimen funnel chart	42
Gambar 4.31 TF-IDF.....	43
Gambar 4.32 Splitting	43
Gambar 4.33 Teknik SMOTE	44
Gambar 4.34 Distribusi model	44

Gambar 4.35 Jumlah sampel dan fitur dalam data latih.....	44
Gambar 4.36 Fungsi fold.....	45
Gambar 4.37 Proses random forest (awal).....	46
Gambar 4.38 Hasil random forest (awal).....	46
Gambar 4.39 Proses random forest (ulang).....	47
Gambar 4.40 Hasil random forest (ulang).....	47
Gambar 4.41 Proses SVM 1 (awal).....	48
Gambar 4.42 Proses SVM 2 (awal).....	48
Gambar 4.43 Hasil SVM (awal).....	49
Gambar 4.44 Proses SVM (ulang).....	49
Gambar 4.45 Hasil SVM (ulang).....	49
Gambar 4.46 Proses XGBoost 1 (awal).....	51
Gambar 4.47 Proses XGBoost 2 (awal).....	51
Gambar 4.48 Hasil XGBoost (awal).....	51
Gambar 4.49 Proses XGBoost 1 (ulang).....	52
Gambar 4.50 Proses XGBoost 2 (ulang).....	52
Gambar 4.51 Hasil XGBoost (ulang).....	52
Gambar 4.52 Proses prediksi 1.....	53
Gambar 4.53 Proses prediksi 2.....	54
Gambar 4.54 Proses prediksi 3.....	54
Gambar 4.55 Hasil prediksi Random Forest.....	54
Gambar 4.56 Hasil prediksi SVM.....	55
Gambar 4.57 Hasil prediksi XGBoost.....	55
Gambar 4.58 Distribusi sentimen acara TV.....	57
Gambar 4.59 Proses visualisasi distribusi sentimen acara TV.....	57
Gambar 4.60 Visualisasi distribusi sentimen acara TV.....	58
Gambar 4.61 Proses visualisasi kata yang sering muncul.....	59
Gambar 4.62 Visualisasi kata yang sering muncul.....	60
Gambar 4.63 Acara TV paling banyak.....	61
Gambar 4.64 Visualisasi Acara TV paling banyak.....	62

DAFTAR TABEL

Tabel 3.1 Pembagian tugas	23
---------------------------------	----

DAFTAR LAMPIRAN

Lampiran 1. Codingan Capstone Project.....	69
Lampiran 2. Load data mentah	70
Lampiran 3. Proses merge data	70
Lampiran 4. Dokumentasi Final Presentasi.....	71