

**HUBUNGAN SENTIMEN DEBAT PILPRES INDONESIA TERHADAP
ELEKTABILITAS PASLON MENGGUNAKAN ALGORITMA NAIVE
BAYES DAN KORELASI PEARSON**



SKRIPSI

**Arnoldus Yitzhak Petra Manoppo
00000048162**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024**

**HUBUNGAN SENTIMEN DEBAT PILPRES INDONESIA TERHADAP
ELEKTABILITAS PASLON MENGGUNAKAN ALGORITMA NAIVE
BAYES DAN KORELASI PEARSON**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Arnoldus Yitzhak Petra Manoppo

00000048162

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Arnoldus Yitzhak Petra Manoppo
Nomor Induk Mahasiswa : 00000048162
Program Studi : Informatika

Skripsi dengan judul:

Hubungan Sentimen Debat Pilpres Indonesia terhadap Elektabilitas Paslon Menggunakan Algoritma Naive Bayes dan Korelasi Pearson

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 21 Mei 2024



(Arnoldus Yitzhak Petra Manoppo)

U M M N
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

HUBUNGAN SENTIMEN DEBAT PILPRES INDONESIA TERHADAP ELEKTABILITAS PASLON MENGGUNAKAN ALGORITMA NAIVE BAYES DAN KORELASI PEARSON

oleh

Nama : Arnoldus Yitzhak Petra Manoppo
NIM : 00000048162
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika


Telah diujikan pada hari Selasa, 4 Juni 2024

Pukul 08.00 s/d 10.00 dan dinyatakan


LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang


14 Juni 2024
(Arya Wicaksana, S.Kom., M.Eng.Sc.
(OCA, CEH, CEI))
NIDN: 0315109103


Penguji


(SY Yuliani Yakub, S.Kom., M.T., Ph.D.)
NIDN: 0411037904

Pembimbing


(Wirawan Istiono, S.Kom., M.Kom.)
NIDN: 0313048304

Pjs. Ketua Program Studi Informatika,


(Dr. Eng. Niki Prastomo, S.T., M.Sc.)
NIDN: 0419128203

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Arnoldus Yitzhak Petra Manoppo

NIM : 00000048162

Program Studi : Informatika

Jenjang : S1

Jenis Karya : Skripsi

Menyatakan dengan sesungguhnya bahwa:

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)**.

Tangerang, 21 Mei 2024

Yang menyatakan



Arnoldus Yitzhak Petra Manoppo

U M M N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

** Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

”Apa pun juga yang kamu perbuat, perbuatlah dengan segenap hatimu seperti untuk Tuhan dan bukan untuk manusia.”

Kolose 3:23 (TB)



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Hubungan Sentimen Debat Pilpres Indonesia terhadap Elektabilitas Paslon Menggunakan Algoritma Naive Bayes dan Korelasi Pearson dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika dan Ketua Program Studi Informatika Universitas Multimedia Nusantara.
3. Bapak Wirawan Istiono, S.Kom., M.Kom., sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 21 Mei 2024



Arnoldus Yitzhak Petra Manoppo

U M N N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

**HUBUNGAN SENTIMEN DEBAT PILPRES INDONESIA TERHADAP
ELEKTABILITAS PASLON MENGGUNAKAN ALGORITMA NAIVE
BAYES DAN KORELASI PEARSON**

Arnoldus Yitzhak Petra Manoppo

ABSTRAK

Kampanye merupakan kegiatan yang berpengaruh terhadap elektabilitas paslon. Debat pilpres merupakan salah satu jenis kampanye untuk memperkenalkan para paslon terhadap masyarakat. Debat pilpres yang ditayangkan pada media sosial Youtube menghasilkan banyak komentar. Untuk mengetahui sentimen setiap komentar secara otomatis dan cepat, dilakukan implementasi *machine learning*. Penelitian ini menggunakan VADER untuk melakukan data *labelling* secara otomatis dan algoritma Naive Bayes untuk melakukan analisis sentimen komentar debat pilpres. Setelah sentimen didapatkan dilakukan perhitungan dan uji korelasi Pearson untuk melihat hubungan antara sentimen debat pilpres pada media sosial terhadap elektabilitas paslon. Dari penelitian yang dilakukan, model Naive Bayes untuk melakukan klasifikasi pada dataset dengan *class* netral memiliki rata-rata akurasi 65%, *precision* 59%, *recall* 57%, dan *F1 score* 57%. Untuk dataset tanpa *class* netral, didapatkan rata-rata akurasi 77%, *precision* 68%, *recall* 68%, dan *F1 score* 67%. Sentimen positif memiliki tingkat hubungan sedang arah negatif sedangkan sentimen negatif memiliki tingkat hubungan kuat arah positif. Baik sentimen positif maupun negatif secara signifikan berhubungan dengan elektabilitas paslon.

Kata kunci: Analisis Sentimen, Debat Pilpres, Korelasi Pearson, *Machine Learning*, Naive Bayes



The Relationship between Sentiments in the Indonesian Presidential Debate and Candidates Electability Using Naive Bayes Algorithm and Pearson Correlation

Arnoldus Yitzhak Petra Manoppo

ABSTRACT

Campaign is an activity that has an impact on the electability of the candidates. The presidential debate is one type of campaign to introduce the candidates to the public. Presidential debates broadcast on social media platforms like YouTube generate a large number of comments. To automatically and quickly determine the sentiment of each comment, a machine learning approach is implemented. This study utilizes VADER to label data automatically and the Naive Bayes algorithm to perform sentiment analysis on presidential debate comments. Once the sentiment is obtained, Pearson correlation is calculated and tested to determine the relationship between the sentiment of presidential debates on social media and candidate electability. The Naive Bayes model for classifying the dataset with a neutral class has an average accuracy of 65%, precision of 59%, recall of 57%, and F1 score of 57%. For the dataset without a neutral class, an average accuracy of 77%, precision of 68%, recall of 68%, and F1 score of 67% is obtained. Positive sentiment has a moderate negative correlation, while negative sentiment has a strong positive correlation. Both positive and negative sentiment are significantly correlated with candidate electability.

Keywords: *Machine Learning, Naive Bayes, Pearson Correlation, Presidential Debate, Sentiment Analysis*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR KODE	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	6
2.1 Pemilu	6
2.2 Debat Pilpres	6
2.3 Natural Language Processing	6
2.4 Analisis Sentimen	7
2.5 Data Preprocess	7
2.6 Valence Aware Dictionary and Sentiment Reasoner	8
2.7 Bag of Words	9
2.8 Synthetic Minority Oversampling Technique	9
2.9 Naive Bayes	10
2.10 Confusion Matrix	10
2.10.1 Accuracy	11
2.10.2 Precision	11
2.10.3 Recall	11
2.10.4 F1 Score	11
2.11 Pearson Correlation Coefficient	11
2.12 Uji t	12
BAB 3 METODOLOGI PENELITIAN	14
3.1 Metodologi Penelitian	14
3.1.1 Telaah Literatur	14
3.1.2 Teknik Pengumpulan Data	14
3.1.3 Perancangan Sistem	17
3.1.4 Evaluasi Model	21
3.1.5 Pehitungan Korelasi	21
3.1.6 Uji Hipotesis	22
BAB 4 HASIL DAN DISKUSI	23
4.1 Pengumpulan Data	23
4.2 Perancangan Sistem	24

4.2.1	Data Preprocessing	24
4.2.2	Implementasi Naive Bayes	31
4.3	Evaluasi Model	35
4.3.1	Evaluasi Model pada Dataset dengan Class Netral	35
4.3.2	Evaluasi Model pada Dataset tanpa Class Netral	40
4.4	Perhitungan Korelasi dan Uji Hipotesis	43
4.4.1	Sentimen Positif	44
4.4.2	Sentimen Negatif	45
BAB 5	SIMPULAN DAN SARAN	47
5.1	Simpulan	47
5.2	Saran	47
	DAFTAR PUSTAKA	48



DAFTAR TABEL

Tabel 1.1	Penelitian terkait	2
Tabel 2.1	Tabel confusion matrix	10
Tabel 2.2	Pedoman derajat korelasi Pearson	12
Tabel 4.1	Hasil text cleaning	25
Tabel 4.2	Hasil case folding	26
Tabel 4.3	Hasil normalize	27
Tabel 4.4	Hasil stopword removal	28
Tabel 4.5	Hasil tokenizing	28
Tabel 4.6	Hasil stemming	29
Tabel 4.7	Hasil terjemahan	30
Tabel 4.8	Hasil data labelling	31
Tabel 4.10	Hasil evaluasi per paslon pada dataset dengan class netral	39
Tabel 4.12	Hasil evaluasi per paslon pada dataset tanpa class netral	43



DAFTAR GAMBAR

Gambar 2.1	Flowchart VADER	8
Gambar 2.2	Flowchart SMOTE	9
Gambar 3.1	Flowchart metodologi	14
Gambar 3.2	Flowchart pengumpulan data	15
Gambar 3.3	Flowchart perancangan sistem	17
Gambar 3.4	Flowchart data preprocessing	18
Gambar 3.5	Flowchart implementasi Naive Bayes	20
Gambar 3.6	Flowchart evaluasi	21
Gambar 3.7	Flowchart perhitungan korelasi	21
Gambar 3.8	Flowchart uji hipotesis	22
Gambar 4.1	Jumlah data	31
Gambar 4.2	Jumlah data tes dan latih	32
Gambar 4.3	Sebelum dan sesudah SMOTE pada dataset dengan class Netral	34
Gambar 4.4	Sebelum dan sesudah SMOTE pada dataset tanpa class netral	34
Gambar 4.5	Confusion matrix pada dataset dengan class netral	36
Gambar 4.6	Classification report pada dataset dengan class netral	37
Gambar 4.7	Confusion matrix pada dataset tanpa class netral	40
Gambar 4.8	Classification report pada dataset tanpa class netral	41



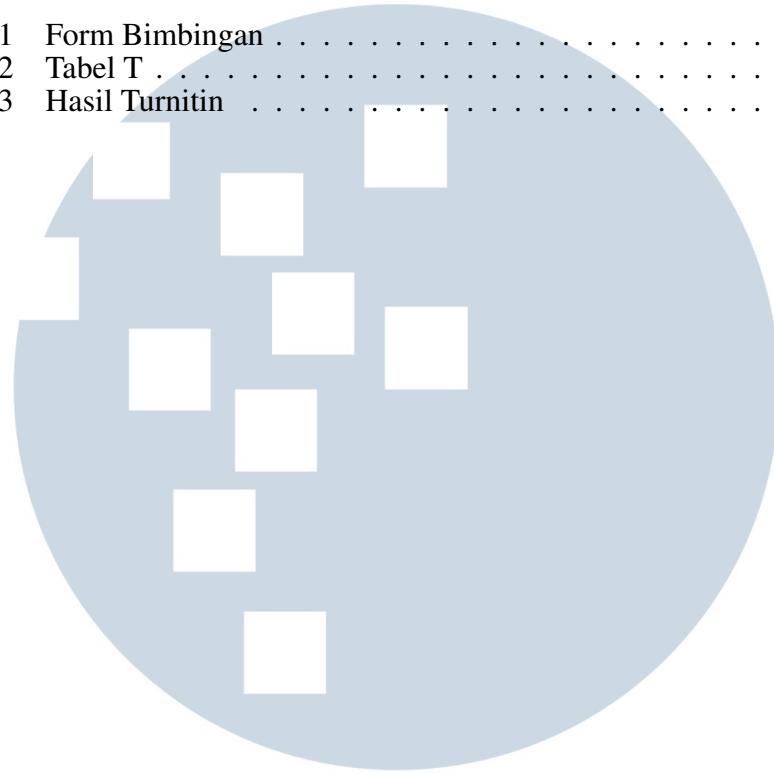
DAFTAR KODE

Kode 4.1	Kode pengambilan data komentar Youtube	23
Kode 4.2	Kode filter berdasarkan keywords	23
Kode 4.3	Kode filter berdasarkan waktu, jumlah kata, dan simpan ke dalam csv.....	24
Kode 4.4	Kode text cleaning.....	25
Kode 4.5	Kode case folding	26
Kode 4.6	Kode normalize	26
Kode 4.7	Kode stopword removal.....	27
Kode 4.8	Kode tokenizing.....	28
Kode 4.9	Kode stemming.....	29
Kode 4.10	Kode data labelling.....	29
Kode 4.11	Kode pembuatan variabel tanpa class netral	31
Kode 4.12	Kode pembagian data latih dan tes	32
Kode 4.13	Kode feature extraction	33
Kode 4.14	Kode oversampling.....	33
Kode 4.15	Kode pelatihan model.....	35
Kode 4.16	Kode evaluasi model	35



DAFTAR LAMPIRAN

Lampiran 1	Form Bimbingan	51
Lampiran 2	Tabel T	52
Lampiran 3	Hasil Turnitin	53



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA