

**ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL
TERHADAP PEMILIHAN PRESIDEN 2024 DENGAN
METODE NAIVE BAYES CLASSIFIER**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ARDHI WAHYU NUGROHO

20.11.3676

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL
TERHADAP PEMILIHAN PRESIDEN 2024 DENGAN
METODE NAIVE BAYES CLASSIFIER**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ARDHI WAHYU NUGROHO

20.11.3676

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

HALAMAN PERSETUJUAN
SKRIPSI
ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TERHADAP PEMILIHAN
PRESIDEN 2024 DENGAN METODE NAIVE BAYES CLASSIFIER

yang disusun dan diajukan oleh

Ardhi Wahyu Nugroho
20.11.3676

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 17 Oktober 2024

Dosen Pembimbing,



Rumini, M.Kom
NIK. 190302246

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TERHADAP
PEMILIHAN PRESIDEN 2024 DENGAN METODE NAIVE BAYES
CLASSIFIER**

yang disusun dan diajukan oleh

Ardhi Wahyu Nugroho

20.11.3676

Telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Oktober 2024

Susunan Dewan Penguji

Nama Penguji

Anna Baita, M.Kom
NIK. 190302290

Dina Maulina, M.Kom
NIK. 190302250

Rumini, M.Kom
NIK. 190302246

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 Oktober 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ardhi Wahyu Nugroho
NIM : 20.11.3676

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Pengguna Media Sosial Terhadap Pemilihan Presiden
2024 Dengan Metode Naive Bayes Classifier**

Dosen Pembimbing : Rumini, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 17 Oktober 2024

Yang Menyatakan,


METRAJ
TEMPEL
E7AMX022025356

Ardhi Wahyu Nugroho

HALAMAN PERSEMBAHAN

Pertama-tama, saya panjatkan puji syukur kepada Allah SWT yang telah memberikan rahmat, taufik, dan hidayah-Nya sehingga saya dapat menyelesaikan skripsi ini dengan baik. Dengan penuh rasa syukur, saya persembahkan halaman persembahan ini sebagai ungkapan terima kasih kepada semua pihak yang telah memberikan dukungan, bimbingan, dan inspirasi selama penulisan skripsi, yang merupakan bagian dari penyelesaian masa studi saya pada Program Studi S1-Informatika di Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Terima kasih ini saya persembahkan dengan tulus kepada:

1. Allah SWT karena hanya atas izin dan karunia-Nya maka skripsi ini dapat di buat dan selesai pada waktunya.
2. Kedua orang tua saya yang telah berkorban banyak tenaga, pikiran, dan materi tanpa lelah dan mengeluh sedikitpun. Semoga Allah SWT selalu memberikan kesehatan dan umur panjang. Amin.
3. Adik satu-satunya yang paling ku cinta, yang mungkin saat saya menyusun skripsi kamu belum paham apa-apa, tapi semoga kamu kelak nanti bisa membacanya ketika kamu sudah memahami semuanya. Tetaplah menjadi adik yang membanggakan.
4. Dosen pembimbing yang telah dengan sabar, bijaksana, dan penuh dedikasi membimbing saya selama proses penyusunan skripsi ini. Terima kasih atas setiap waktu, tenaga, dan perhatian yang Ibu berikan. Semoga segala kebaikan Ibu mendapatkan balasan yang setimpal dari Tuhan Yang Maha Esa.
5. Teman – teman Universitas Amikom Yogyakarta yang telah berbagi pengetahuan, pengalaman satu sama lain selama perjalanan akademik kami di perguruan tinggi.

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, petunjuk, dan karunia-Nya yang telah membantu penulis menyelesaikan skripsi dengan judul “Analisis Sentimen Pengguna Media Sosial Terhadap Pemilihan Presiden 2024 Dengan Metode Naive Bayes Classifier.” Skripsi ini disusun untuk memenuhi persyaratan kelulusan Program Studi S1-Informatika pada Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Penulis menyadari bahwa penyusunan skripsi ini tidak terlepas dari doa, bimbingan, dukungan, dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Keluarga tersayang, yang senantiasa memberikan doa terbaik, dukungan moral, materi, serta semangat dalam perjalanan penyusunan skripsi ini.
2. Prof. Dr. M. Suyanto., M.M selaku Rektor Universitas Amikom Yogyakarta
3. Rumini, M.Kom selaku dosen pembimbing yang telah memberikan bimbingan dan berbagai pengalaman kepada penulis
4. Segenap Dosen Program Studi Informatika yang telah mendidik serta memberikan ilmu selama kuliah dan seluruh staff yang selalu sabar melayani segala administrasi
5. Sahabat dan teman, yang telah berbagi ilmu, pengalaman, dan inspirasi sepanjang perjalanan akademik di Universitas Amikom Yogyakarta.
6. Semua individu yang turut menyumbangkan gagasan, pandangan, dan sokongan dalam berbagai bentuk, yang telah membantu kelancaran perjalanan penyelesaian skripsi ini

Yogyakarta, 17 Oktober 2024

Ardhi Wahyu Nugroho

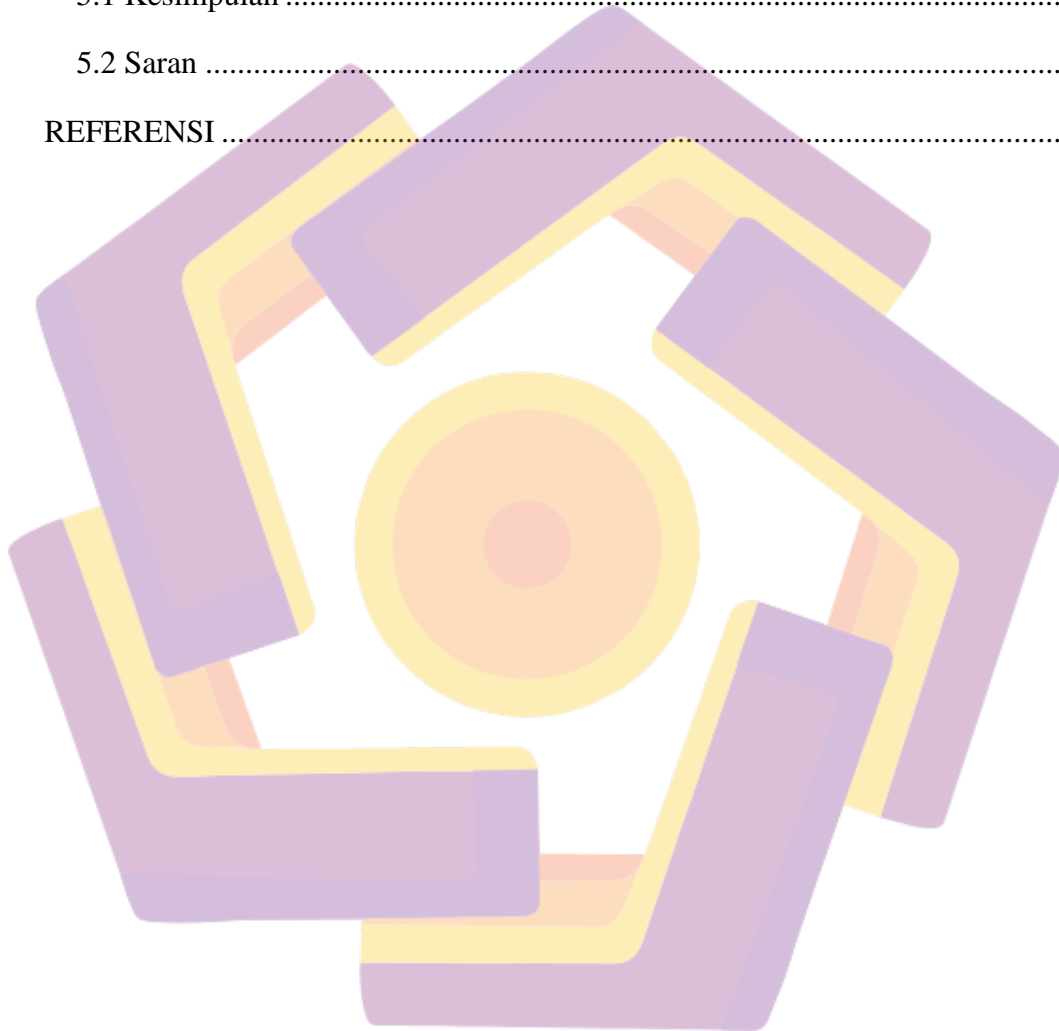
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iii
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xv
INTISARI	xvii
<i>ABSTRACT</i>	xviii
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	5
1.5.1 Manfaat Teoritis.....	5
1.5.2 Manfaat Praktis	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	8

2.1 Studi Literatur	8
2.2 Dasar Teori	21
2.2.1 Analisis Sentimen	21
2.2.2 Crawling Data	21
2.2.3 Text Mining	21
2.2.4 Data Preprocessing.....	22
2.2.4.1 Pembersihan Data (Data Cleaning).....	22
2.2.4.2 Case Folding	22
2.2.4.3 Remove Stopword.....	22
2.2.4.4 Tokenization	23
2.2.4.5 Stemming	23
2.2.4.6 Normalisasi	24
2.2.5 Labeling	24
2.2.6 Bag of Words (BoW)	24
2.2.7 TF-IDF (Term Frequence Inverse Document Frequency)	25
2.2.8 Algoritma Naïve Bayes	25
BAB III METODE PENELITIAN	29
3.1 Alur Penelitian	29
3.1.1 Study literatur.....	30
3.1.2 Pengumpulan Data	30
3.1.3 Data Preprocessing.....	31
3.1.3.1 Data Cleaning.....	31
3.1.3.2 Normalize.....	31
3.1.3.3 Remove Stopword.....	32
3.1.3.4 Tokenization	33

3.1.3.1 Stemming	34
3.2.4 Labeling	35
3.2.5 Visualisasi	36
3.2.6 Pembagian Data	36
3.2.7 Feature Ekstraktion	37
3.2.8 Ekstraksi Hasil	38
3.2.9 Evaluasi	39
BAB IV HASIL DAN PEMBAHASAN	41
4.1 Dataset Crawling.....	41
4.2 Preprocessing Data.....	42
4.2.1 Cleaning Dataset	45
4.2.2 Normalize.....	45
4.2.3 Remove Stopword.....	47
4.2.4 Tokenize	48
4.2.5 Stemming	49
4.3 Labeling	50
4.4 Visualisasi	52
4.5 Pembagian data	52
4.6 Feature Engineering	53
4.6.1 Bag of Words (BOW)	53
4.6.2 TF-IDF	55
4.7 Ekstraksi Hasil	56
4.7.1 Bag of Words (BOW)	56
4.7.2 TF-IDF	57
4.8 Evaluasi.....	58

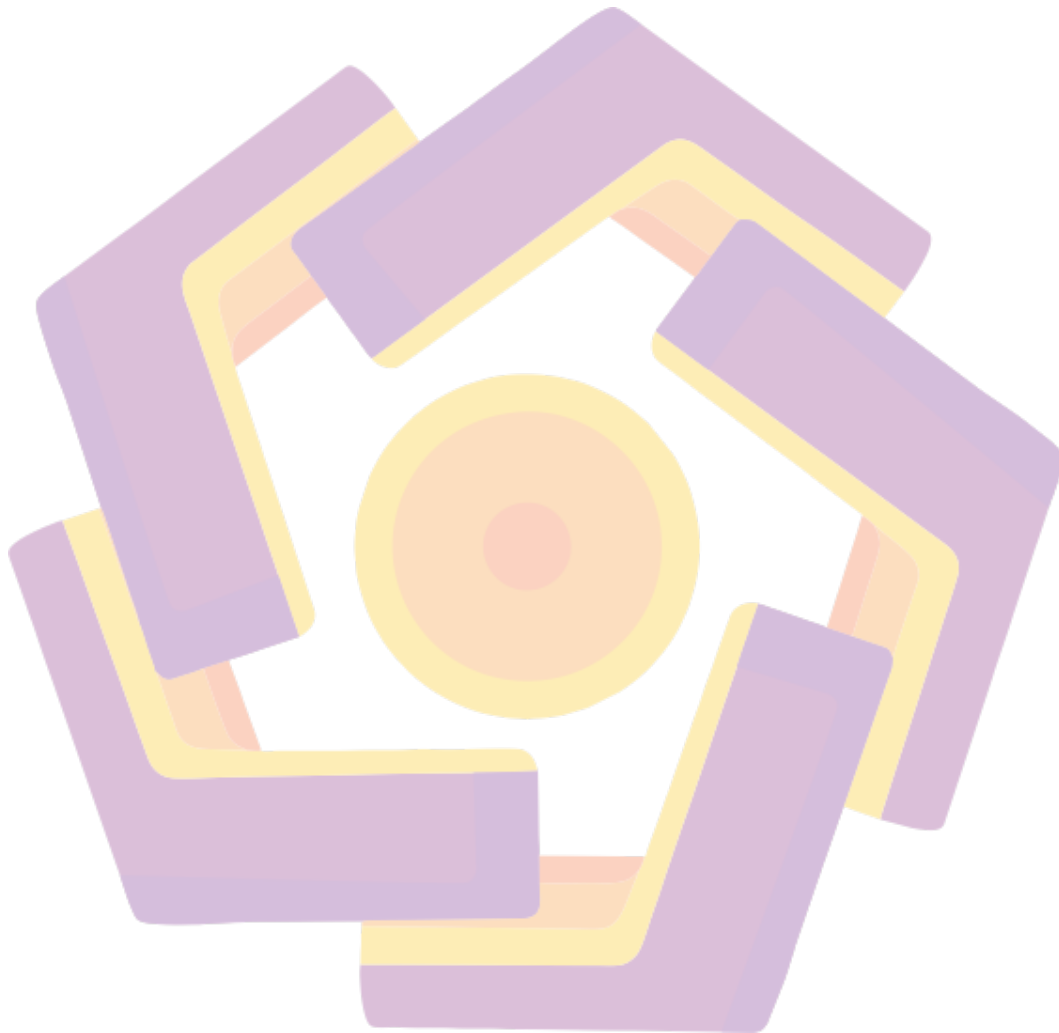
4.8.1 Classification report Bow dan TF-IDF	58
4.8.2 Confusion matrix Model Bow	59
4.8.3 Confusion matrix Model TF-IDF.....	60
BAB V PENUTUP	61
5.1 Kesimpulan	61
5.2 Saran	61
REFERENSI	63



DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian12

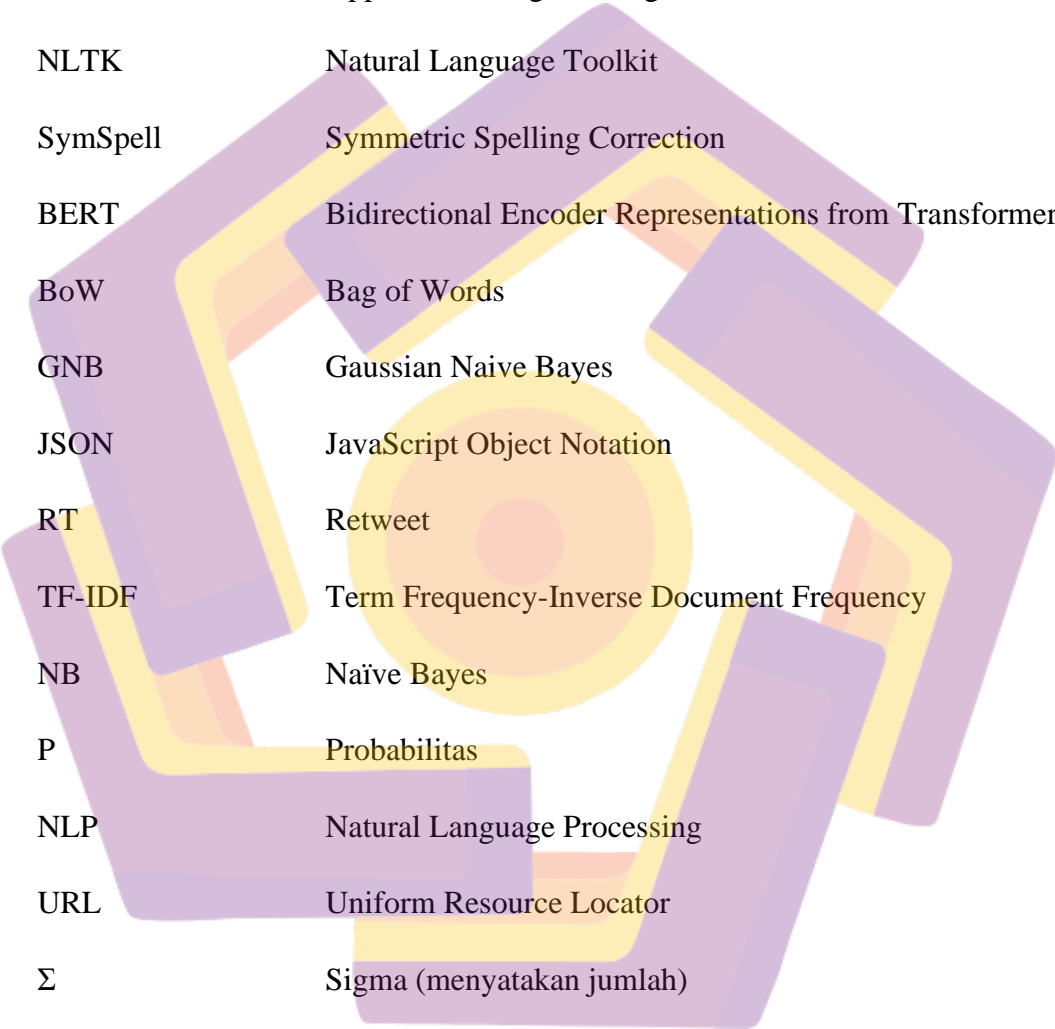
Tabel 4. 1 Hasil Evaluasi Dua Model Naive Bayes Yaitu Bow Dan Tf-Idf58



DAFTAR GAMBAR

Gambar 2. 1 Naive Bayes	26
Gambar 3. 1 Alur Penelitian	29
Gambar 4. 1 Proses Crawl Data	41
Gambar 4. 2 hasil pembacaan file csv	42
Gambar 4. 3 Preprocessing Data Dan Penghapusan Kolom.....	43
Gambar 4. 4 Pengecekan Data Isnull(Nan).....	44
Gambar 4. 5 Setelah Penghapusan Nilai Nan	44
Gambar 4. 6 Hasil Cleaning Dataset.....	45
Gambar 4. 7 Hasil Normalize	46
Gambar 4. 8 Hasil Remove Stopword	47
Gambar 4. 9 Hasil Tokenize	48
Gambar 4. 10 Hasil Proses Stemming	49
Gambar 4. 11 Hasil Labeling Dataset	50
Gambar 4. 12 Hasil Proses Visualisasi Dataset	52
Gambar 4. 13 Hasil Proses Pembagian Dataset	53
Gambar 4. 14 Proses Feature Engineering Menggunakan Bag Of Words (Bow) ...	54
Gambar 4. 15 Proses Feature Engineering Tf-Idf.....	55
Gambar 4. 16 Proses Ekstraksi Hasil Menggunakan Bag Of Words (Bow)	56
Gambar 4. 17 Proses Ekstraksi Hasil Menggunakan Tf-Idf	57
Gambar 4. 18 Confusion Matrix Model Bow	59
Gambar 4. 19 Confusion Matrix Model Tf-Idf	60

DAFTAR LAMBANG DAN SINGKATAN



NLP	Natural Language Processing
X	Twitter
API	Application Programming Interface
NLTK	Natural Language Toolkit
SymSpell	Symmetric Spelling Correction
BERT	Bidirectional Encoder Representations from Transformers
BoW	Bag of Words
GNB	Gaussian Naive Bayes
JSON	JavaScript Object Notation
RT	Retweet
TF-IDF	Term Frequency-Inverse Document Frequency
NB	Naïve Bayes
P	Probabilitas
NLP	Natural Language Processing
URL	Uniform Resource Locator
Σ	Sigma (menyatakan jumlah)
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
F1	F1 Score

V_j	kategori tweet $j = 1, 2, 3, \dots, n$. Dimana dalam penelitian ini
j_1	Kategori tweet sentiment negative.
j_2	kategori tweet sentiment positif.
j_3	kategori tweet sentiment netral.
$P(x_i V_j)$	Probabilitas x_i pada kategori
$V_j P(V_j)$	Probabilitas dari V_j .
$ docs_j $	jumlah dokumen setiap kategori j
$ contoh $	jumlah dokumen dari semua kategori
n_k	jumlah frekuensi kemunculan setiap kata
n	jumlah frekuensi kemunculan kata dari setiap kategori
$ Kosakata $	jumlah semua kata dari semua kategori

DAFTAR ISTILAH

Analisis Sentimen	Menentukan apakah teks memiliki sentimen positif, negatif, atau netral.
Algoritma Naive Bayes	Sebuah algoritma klasifikasi yang menggunakan teorema Bayes untuk menentukan probabilitas tertinggi dari suatu data, sering digunakan dalam analisis sentimen.
Crawling	Proses mengumpulkan informasi dari internet.
Dataset	Kumpulan data yang digunakan untuk analisis atau penelitian
Google Colab	Platform cloud untuk pemrograman Python dan analisis data
Preprocessing	Proses persiapan data sebelum dilakukan analisis atau pemodelan, termasuk pembersihan dan normalisasi data.
Akurasi	Tingkat ketepatan model dalam mengklasifikasikan data ke dalam kategori yang benar.
Presisi	Ukuran sejauh mana hasil yang diperoleh dari model klasifikasi merupakan hasil yang relevan
Recall	Ukuran kemampuan model untuk mengidentifikasi semua data relevan dalam dataset.
Visualisasi Data	Representasi grafis dari data untuk membantu dalam memahami dan menganalisis informasi yang terkandung di dalamnya.
NLTK	Library Python untuk pemrosesan bahasa alami.
Node.js	Runtime JavaScript untuk menjalankan kode di luar browser
Normalisasi	Mengubah kata-kata ke bentuk standar.
Pandas	Library Python untuk analisis dan manipulasi data tabular
Sastrawi	Library bahasa Indonesia untuk pemrosesan teks
Seaborn	Library Python untuk visualisasi data statistik
Splitting	Pembagian dataset menjadi dua bagian.
Stemming	Menghilangkan infleksi kata.
Stopwords Removal	Menghapus kata-kata umum dari teks.

Text Mining	Ekstraksi informasi dari teks yang tidak terstruktur menggunakan metode dan teknik tertentu.
Tweet-harvest	Pengumpulan tweet dari Twitter.
Word Cloud	Representasi visual frekuensi kata-kata dalam teks
Case Folding	Proses mengubah kata menjadi huruf kecil untuk menjaga konsistensi dalam data teks.
Tokenization	Proses memecah teks menjadi serangkaian unit unik yang disebut token.
Labeling	Proses memberikan label atau kategori pada dataset untuk mengidentifikasi sentimen di dalamnya.
Bag of Words (BoW)	Model representasi teks sebagai multiset dari kata-kata tanpa mempertimbangkan urutan kata.
TF-IDF	Perhitungan bobot pada tiap kata dalam dokumen untuk menilai relevansi kata terhadap dokumen.
Probabilitas Prior	Peluang dari setiap kategori sebelum mengamati data.
Pengujian Performa	Evaluasi terhadap akurasi, presisi, dan recall dari algoritma yang diterapkan.
F1-Score	Rata-rata harmonik dari presisi dan recall, memberikan gambaran keseluruhan tentang kinerja model.

INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Pemilihan Presiden 2024 melalui media sosial, khususnya Twitter. Dengan semakin berkembangnya penggunaan media sosial sebagai sarana komunikasi publik, banyak opini, komentar, dan sentimen yang disampaikan secara terbuka. Namun, volume data yang besar serta penggunaan bahasa informal di platform ini menimbulkan kesulitan dalam memahami sentimen secara efektif. Dampak dari permasalahan ini adalah kurangnya pemahaman yang akurat tentang preferensi politik masyarakat, yang dapat memengaruhi strategi kampanye dan kebijakan publik. Oleh karena itu, diperlukan sebuah metode untuk mengklasifikasikan sentimen secara lebih efektif dari data yang tidak terstruktur. Untuk menyelesaikan masalah tersebut, penelitian ini menggunakan algoritma Naive Bayes Classifier untuk mengklasifikasikan sentimen tweet menjadi tiga kategori: positif, netral, dan negatif. Data yang digunakan dikumpulkan dari komentar tweet terkait Pilpres 2024 dengan kata kunci tertentu. Analisis dilakukan menggunakan Google Colaboratory dengan bahasa pemrograman Python. Algoritma Naive Bayes dipilih karena kemampuannya yang efisien dalam menangani data dalam jumlah besar dan komputasi yang sederhana. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memiliki akurasi yang baik dalam mengklasifikasikan sentimen tweet. Kontribusi penelitian ini adalah memberikan wawasan tentang pandangan masyarakat terhadap calon presiden, yang dapat dimanfaatkan oleh partai politik, peneliti, atau pengamat sosial untuk memahami dinamika opini publik. Diharapkan Penelitian lebih lanjut dapat berfokus pada pengembangan model analisis sentimen yang lebih akurat dengan mempertimbangkan penggunaan algoritma lain.

Kata kunci: Analisis Sentimen, Naive Bayes Classifier, Twitter, Pemilihan Presiden 2024, Media Sosial

ABSTRACT

This study aims to analyze public sentiment towards the 2024 Presidential Election through social media, especially Twitter. With the increasing use of social media as a means of public communication, many opinions, comments, and sentiments are conveyed openly. However, the large volume of data and the use of informal language on this platform make it difficult to understand sentiment effectively. The impact of this problem is the lack of accurate understanding of people's political preferences, which can affect campaign strategies and public policies. Therefore, a method is needed to classify sentiment more effectively from unstructured data. To solve this problem, this study uses the Naive Bayes Classifier algorithm to classify tweet sentiment into three categories: positive, neutral, and negative. The data used was collected from tweet comments related to the 2024 Presidential Election with certain keywords. The analysis was carried out using Google Colaboratory with the Python programming language. The Naive Bayes algorithm was chosen because of its efficient ability to handle large amounts of data and simple computation. The results of the study show that the Naive Bayes algorithm has good accuracy in classifying tweet sentiment. The contribution of this research is to provide insight into the public's views on presidential candidates, which can be utilized by political parties, researchers, or social observers to understand the dynamics of public opinion. It is hoped that further research can focus on developing a more accurate sentiment analysis model by considering the use of other algorithms.

Keyword: *Sentiment Analysis, Naive Bayes Classifier, Twitter, 2024 Presidential Election, SocialMedia*