

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR ISTILAH	xv
DAFTAR LAMPIRAN.....	xix
ABSTRAK	xx
ABSTRACT.....	xxi
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Rumusan Masalah.....	5
C. Batasan Masalah	6
D. Tujuan Penelitian	6
E. Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA.....	9
A. Landasan Teori.....	9
B. Penelitian Sebelumnya.....	24
BAB III METODE PENELITIAN.....	32
A. Metode Pengumpulan Data.....	32
B. Alat dan Bahan Penelitian.....	33
C. Konsep Penelitian	34
BAB IV HASIL DAN PEMBAHASAN	46

A. Hasil.....	46
B. Pembahasan	52
BAB V PENUTUP.....	60
A. Kesimpulan	60
B. Saran	63

DAFTAR PUSTAKA

LAMPIRAN



DAFTAR TABEL

Tabel 2. 1 Perbandingan Penelitian terdahulu	29
Tabel 3. 1 Alat dan Bahan Penelitian.....	33
Tabel 3. 2 Skenario Eksperimen	40
Tabel 4. 1 Kombinasi 2 representasi fitur	47
Tabel 4. 2 Hasil Pengujian	48
Tabel 4. 3 Top 20 Kata dengan Bobot Tertinggi per Kelas untuk TF-IDF.....	51
Tabel 5. 1 Ringkasan Perbandingan Kinerja Tiga Algoritma Klasifikasi.....	62



DAFTAR GAMBAR

Gambar 3. 1 Kerangka Berpikir	36
Gambar 4. 1 Confusion Matrix Logistic Regression + BERT	49
Gambar 4. 2 Confusion Matrix Naïve Bayes(Gaussian) + BERT	50
Gambar 4. 3 Ringkasan Hasil Demo.....	58



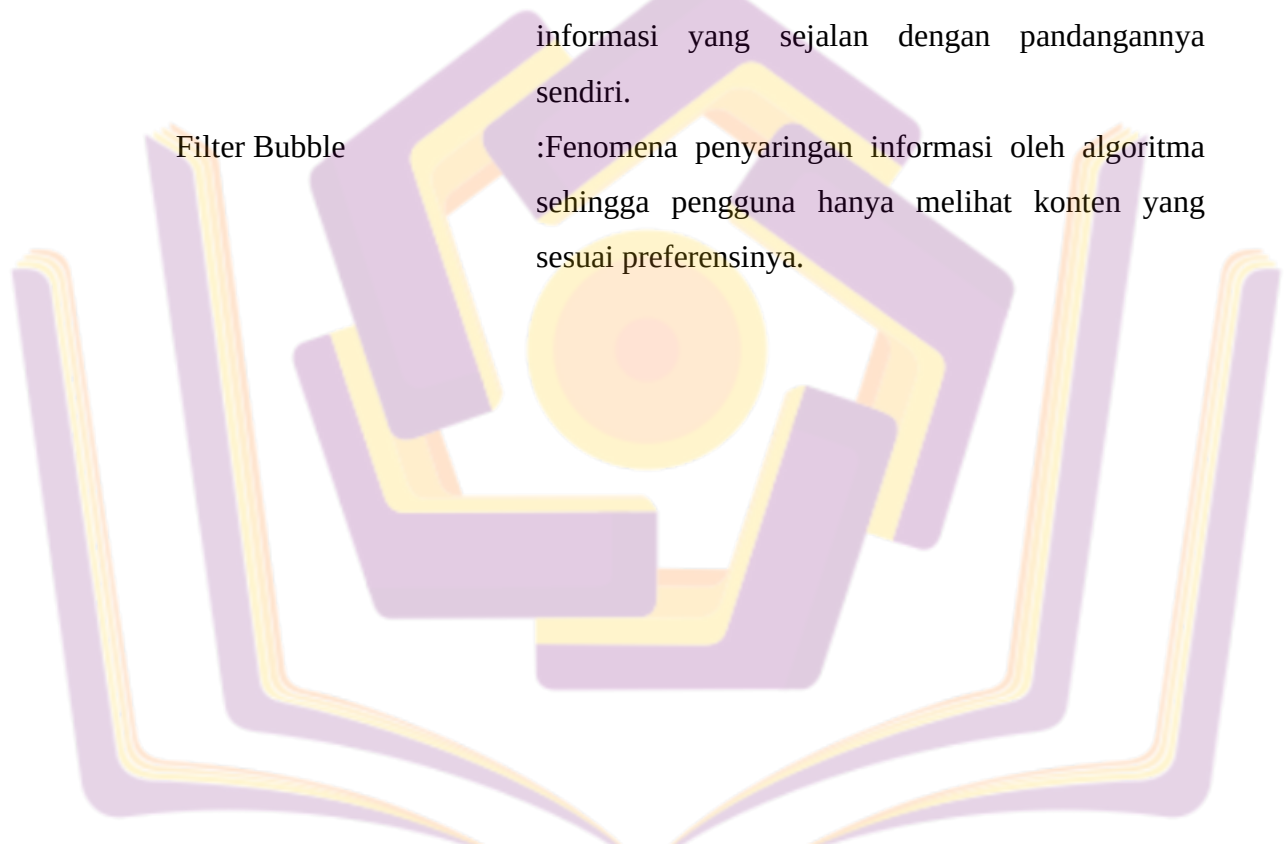
DAFTAR ISTILAH

Hoaks	:Informasi yang tidak sesuai dengan fakta, tetapi dikemas seolah-olah benar untuk memengaruhi persepsi publik.
Misinformation	:Informasi yang salah tanpa adanya niat untuk menimbulkan kerugian.
Disinformation	:Informasi palsu yang sengaja dibuat untuk menyesatkan atau merugikan pihak tertentu.
Mal-information	:Informasi yang berdasarkan fakta tetapi digunakan untuk merugikan individu, organisasi, atau kelompok tertentu.
Text Mining	:Proses untuk mengekstraksi informasi dan pengetahuan yang bermanfaat dari kumpulan dokumen teks dengan memanfaatkan teknik data mining, NLP, dan machine learning.
Preprocessing	:Proses pembersihan dan penyederhanaan data teks agar lebih mudah diproses oleh algoritma machine learning.
Case Folding	:Proses penyeragaman huruf dalam teks, umumnya menjadi huruf kecil semua.
Tokenisasi	:Proses memecah teks menjadi unit-unit kata atau token.
Stopword Removal	:Proses penghapusan kata-kata umum yang tidak memiliki pengaruh signifikan terhadap analisis teks.
TF-IDF (Term Frequency-Inverse Document Frequency)	:Metode representasi fitur yang memberikan bobot pada kata berdasarkan frekuensi kemunculannya dalam dokumen dibandingkan seluruh koleksi dokumen.
Term Frequency (TF)	:Ukuran seberapa sering suatu kata muncul dalam sebuah dokumen.

Inverse Document Frequency (IDF)	:Ukuran yang mengurangi pengaruh kata-kata yang muncul terlalu sering pada seluruh dokumen.
BERT (Bidirectional Encoder Representations from Transformers)	:Model representasi bahasa berbasis Transformer yang mampu memahami konteks kata secara dua arah.
IndoBERT	:Model BERT yang telah dilatih khusus pada korpus teks berbahasa Indonesia.
Contextual Embedding	:Representasi vektor suatu kata yang mempertimbangkan konteks berdasarkan kata-kata di sekitarnya dalam kalimat.
Transformer	:Arsitektur jaringan saraf yang menggunakan mekanisme self-attention untuk memproses urutan teks.
Self-Attention	:Mekanisme dalam Transformer yang mempelajari hubungan antar kata dalam suatu urutan teks.
Fine-tuning	:Proses melatih ulang model pralatih pada dataset spesifik untuk menyesuaikan kemampuannya pada tugas tertentu.
Feature Extraction	:Penggunaan model pralatih hanya sebagai penghasil representasi fitur tanpa melatih ulang parameternya.
Machine Learning	:Metode pembelajaran mesin yang memungkinkan sistem mempelajari pola dari data berlabel dan menggunakannya untuk memprediksi data baru.
Klasifikasi	:Teknik dalam machine learning untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola yang terdapat pada data.
Naive Bayes	:Algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen satu sama lain.
Multinomial Naive Bayes	:Varian <i>Naive Bayes</i> yang cocok untuk data fitur non-negatif seperti TF-IDF.

Gaussian Naive Bayes	: Varian <i>Naive Bayes</i> yang mengasumsikan distribusi normal pada setiap fitur, digunakan untuk representasi BERT.
Logistic Regression	: Metode klasifikasi yang memprediksi probabilitas suatu data termasuk dalam kelas tertentu menggunakan fungsi logistik.
Support Vector Machine (SVM)	: Algoritma klasifikasi yang bekerja dengan mencari hyperplane optimal yang memisahkan data ke dalam dua kelas dengan margin terbesar.
Hyperplane	: Batas pemisah antar kelas pada algoritma SVM.
Confusion Matrix	: Tabel yang menggambarkan kinerja model klasifikasi dengan membandingkan hasil prediksi dengan kondisi sebenarnya.
True Positive (TP)	: Jumlah data hoaks yang diprediksi benar sebagai hoaks.
True Negative (TN)	: Jumlah data fakta yang diprediksi benar sebagai fakta.
False Positive (FP)	: Jumlah data fakta yang salah diprediksi sebagai hoaks.
False Negative (FN)	: Jumlah data hoaks yang salah diprediksi sebagai fakta.
Accuracy	: Proporsi prediksi yang benar terhadap seluruh data yang diuji.
Precision	: Tingkat ketepatan model dalam memprediksi kelas positif.
Recall	: Kemampuan model dalam menemukan seluruh data yang termasuk ke dalam kelas positif.
F1-score	: Rata-rata harmonis antara <i>precision</i> dan <i>recall</i> sebagai ukuran performa model yang seimbang.

Data Leakage	:Kondisi ketika model secara tidak sengaja mempelajari informasi di luar konten berita, seperti pola redaksional sumber media.
Source Artifact Removal	:Tahap pembersihan elemen penanda sumber berita dari teks agar model tidak bias terhadap identitas media.
Natural Language Processing (NLP)	:Cabang ilmu komputer yang berfokus pada interaksi antara komputer dan bahasa manusia.
Echo Chamber	:Kondisi di mana seseorang hanya terpapar informasi yang sejalan dengan pandangannya sendiri.
Filter Bubble	:Fenomena penyaringan informasi oleh algoritma sehingga pengguna hanya melihat konten yang sesuai preferensinya.



DAFTAR LAMPIRAN

- Lampiran 1. Kartu Bimbingan Dosen Pembimbing 1
- Lampiran 2. Kartu Bimbingan Dosen Pembimbing 2
- Lampiran 3. *Confusion Matrix* Seluruh Skenario
- Lampiran 4. Hasil Kinerja Model: Tf-Idf Dan Bert Di Setiap Algoritma
- Lampiran 5. Timeline Pengerjaan Penelitian
- Lampiran 6. Script Kode Algoritma
- Lampiran 7. Contoh Dataset Fakta Sebelum Dibersihkan
- Lampiran 8. Contoh Dataset Hoaks Sebelum Dibersihkan

