

SKRIPSI

**PENERAPAN ALGORITMA NAÏVE BAYES UNTUK
MENDETEKSI BERITA HOAKS DI TWITTER
MENGUNAKAN EKSPANSI FITUR DENGAN GLOVE**



RISKI AMBARWATI

1519620007

PROGRAM STUDI

SISTEM DAN TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS NEGERI JAKARTA

2024


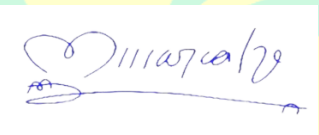
LEMBAR PENGESAHAN SKRIPSI

Judul : Penerapan Algoritma Naïve bayes Untuk Mendeteksi Berita Hoaks di Twitter Menggunakan Ekspansi Fitur dengan GloVe.

Penyusun : Riski Ambarwati

NIM : 1519620007

Skripsi telah didiskusikan dan diusulkan dari Dosen Pembimbing:

NAMA DOSEN	TANDA TANGAN	TANGGAL
Irma Permata Sari, S.Pd.,M.Eng (Dosen Pembimbing I)		10 Juni 2024
Murien Nugraheni, S.T., M.Cs. (Dosen Pembimbing II)		10 Juni 2024

Telah disetujui oleh:

NAMA DOSEN	TANDA TANGAN	TANGGAL
Fuad Mumtas, M.TI (Dosen Penguji 1)		15 Juli 2024

Ali Idrus, S.Kom.,M.Kom.
(Dosen Penguji 2)



15 Juli 2024

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa :

1. Skripsi ini yang berjudul "Penerapan Algoritma Naive Bayes Untuk Mendeteksi Berita Hoaks di Twitter Menggunakan Ekspansi Fitur dengan GloVe" merupakan Karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Skripsi ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 15 Juli 2024

Yang Membuat Pernyataan



Riski Ambarwati

No. Reg. 1519620007

SITAS NEGERI JA



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220

Telepon/Faksimili: 021-4894221

Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA
ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Riski Ambarwati

NIM : 1519620007

Fakultas/Prodi : Sistem dan Teknologi Informasi

Alamat email : riskiambarwati7@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (...)

yang berjudul : Penerapan Algoritma Naïve bayes Untuk Mendeteksi Berita Hoaks di Twitter Menggunakan Ekspansi Fitur dengan GloVe.

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 22 Juli 2024

(Riski Ambarwati)

KATA PENGANTAR

Puji dan syukur penulis ucapkan ke hadirat Allah SWT yang telah melimpahkan rahmat serta karunia-Nya sehingga penulis dapat menyelesaikan Skripsi yang berjudul “Penerapan Algoritma Naïve bayes Untuk Mendeteksi Berita Hoaks di Twitter Menggunakan Ekspansi Fitur dengan GloVe”. Skripsi ini disusun sebagai salah satu persyaratan meraih gelar Sarjana Pendidikan pada Program Studi S1 Sistem dan Teknologi Informasi Universitas Negeri Jakarta. Penyelesaian skripsi ini tidak lepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan ucapan terima kasih kepada :

1. Lipur Sugiyanta, Ph.D, selaku Koordinator Program Studi Sistem dan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Jakarta;
2. Irma Permata Sari selaku Dosen Pembimbing I Proposal Skripsi;
3. Murien Nugraheni selaku Dosen Pembimbing II Proposal Skripsi;
4. Seluruh Dosen Sistem dan Teknologi Informasi Universitas Negeri Jakarta yang telah memberi bekal pengetahuan yang berharga.
5. Teman-teman seperjuangan yang setia saling membantu dalam proses Skripsi.

Penulis hanya dapat memanjatkan doa semoga semua pihak yang telah membantu penulis dalam penyusunan Skripsi ini mendapatkan pahala dari Allah SWT. Semoga Skripsi ini dapat bermanfaat dan memberikan sumbangan yang berarti bagi pihak yang membutuhkan.

Jakarta, 10 Juni 2024



Riski Ambarwati

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	ii
LEMBAR PERNYATAAN.....	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN.....	x
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Identifikasi Masalah.....	3
1.3 Pembatasan Masalah.....	3
1.4 Perumusan Masalah	4
1.5 Tujuan Penelitian	4
1.6 Manfaat Penelitian	4
BAB II.....	5
TINJAUAN PUSTAKA	5
2.1 Kerangka Teoritik	5
2.1.1. Machine Learning	5
2.1.2. Data Mining	6
2.1.3. Text Mining	6
2.1.4. Analisis Sentimen	7
2.1.5. Data Crawling	7
2.1.6. Naïve Bayes Classifier	8
2.1.7. Confusion Matrix	8
2.1.8. Global Vector (GloVe).....	11
2.1.9. Knowledge Discovery of Data (KDD).....	12
2.2 Penelitian Relevan	13
2.3 <i>State of the Art</i>	19
2.4 Kerangka Berpikir.....	19
BAB III	21
METODOLOGI PENELITIAN.....	21
3.1 Tempat dan Waktu Penelitian.....	21
3.2 Metodologi Penelitian.....	21
3.3 Alur Penelitian	21

3.3.1.	Tahap Pengumpulan Data (Data Selection)	21
3.3.2.	Pre Processing	24
3.3.3.	Transformation.....	29
3.3.4.	Data Mining	30
3.3.5.	Interpretation / Evaluation	33
BAB IV	34
HASIL DAN PEMBAHASAN	34
4.1	Hasil Penelitian	34
4.1.1	Hasil Data Selection.....	34
4.1.2	Hasil Pre Processing.....	36
4.1.3	Hasil Transformation Data.....	47
4.1.4	Hasil Data Mining.....	49
4.1.5	Hasil Evaluation.....	50
4.2	Pembahasan.....	60
BAB V	62
KESIMPULAN DAN SARAN	62
5.1	Kesimpulan	62
5.2	Saran	62
DAFTAR PUSTAKA	63
LAMPIRAN	67
DAFTAR RIWAYAT HIDUP	75

DAFTAR TABEL

Tabel 2. 1 Model Confusion Matrix.....	9
Tabel 2. 2 Penelitian Relevan.....	13
Tabel 4. 1 Atribut Pada Dataset	35
Tabel 4. 2 Source Code Cleaning.....	36
Tabel 4. 3 Hasil Dataset Cleaning.....	36
Tabel 4. 4 Souce Code Case Folding	38
Tabel 4. 5 Sampel Hasil Case Folding	39
Tabel 4. 6 Soure Code Tokening.....	40
Tabel 4. 7 Sampel Hasil Tokenizing.....	41
Tabel 4. 8 Source Code Spelling Normalization.....	43
Tabel 4. 9 Sampel Hasil Spelling Normalization.....	43
Tabel 4. 10 Source Code Stopword.....	44
Tabel 4. 11 Sampel Hasil Filtering	44
Tabel 4. 12 Soure Code Stemming	46
Tabel 4. 13 Sampel Hasil Stemming.....	46
Tabel 4. 14 Source Code Embedding Glove Not Clean.....	48
Tabel 4. 15 Source Code Embedding GloVe Clean.....	48
Tabel 4. 16 Perbandingan Hasil Embedding GloVe.....	49
Tabel 4. 17 Tabel Skenario Pembagian Split Data	50
Tabel 4. 18 Source Code AUC.....	50
Tabel 4. 19 Hasil Nilai AUC Setiap Skenario.....	59
Tabel 4. 20 Hasil perbandingan Evaluasi Model Naïve bayes.....	60

DAFTAR GAMBAR

Gambar 1. 1 Pengguna Twitter Lebih Khawatir Berita Hoaks di Internet.....	2
Gambar 2. 1 Tahapan Knowledge Discovery of Database	12
Gambar 2. 2 Kerangka Berpikir	20
Gambar 3. 1 Tahapan Knowledge Discovery of Data (KDD)	21
Gambar 3. 2 Diagram Alur Pengumpulan Data	22
Gambar 3. 3 Install Library Tweet Harvest	22
Gambar 3. 4 Sampel Pengambilan Data Keyword @PemiluDetikcom.....	23
Gambar 3. 5 Sampel Pengambilan Data Pada Keyword @PemiluKompas	23
Gambar 3. 6 Sampel Pengambilan Data Pada Keyword @TurnBackHoax	24
Gambar 3. 7 Diagram Alur Preprocessing	25
Gambar 3. 8 Diagram Alur Proses Cleaning.....	25
Gambar 3. 9 Diagram Alur Tahap Case Folding	26
Gambar 3. 10 Diagram Alur Tahap Tokenizing	27
Gambar 3. 11 Diagram Alur Spelling Normalization	27
Gambar 3. 12 Diagram Proses Filtering.....	28
Gambar 3. 13 Diagram Alur Proses Stemming.....	29
Gambar 4. 1 Sampel Dataset Hasil Crawling.....	34
Gambar 4. 2 Kamus Spelling Normalization	43
Gambar 4. 3 Hasil Sampel Embedding GloVe Not Clean	48
Gambar 4. 4 Hasil Sampel Embedding GloVe Clean	49
Gambar 4. 5 Library Gaussian Bayes	49
Gambar 4. 6 Hasil Skenario 1 Not Clean.....	51
Gambar 4. 7 Hasil Skenario 2 Not Clean.....	52
Gambar 4. 8 Hasil Skenario 3 Not Clean.....	53
Gambar 4. 9 Hasil Skenario 4 Not Clean.....	54
Gambar 4. 10 Hasil Skenario 1 Clean.....	55
Gambar 4. 11 Hasil Skenario 2 Clean.....	56
Gambar 4. 12 Hasil Skenario 3 Clean.....	57
Gambar 4. 13 Hasil Skenario 4 Clean.....	58
Gambar 4. 14 Grafik ROC Hasil Skenario 60:40	59
Gambar 4. 15 Wordcloud keyword 'hoax pemilu 2024'.....	61

DAFTAR LAMPIRAN

Lampiran 1. Surat Tugas Dosen Pembimbing	67
Lampiran 2. Lembar Pernyataan Kelayakan Judul Skripsi	68
Lampiran 3. Surat Permohonan Ujian Skripsi	69
Lampiran 4. Lembar Konsultasi Dosen Pembimbing	70
Lampiran 5. Surat Pernyataan Dosen Pembimbing 1	72
Lampiran 6. Surat Pernyataan Dosen Pembimbing 2	73
Lampiran 7. Lembar Persetujuan Dosen Pembimbing	74

