

**IMPLEMENTASI PRETRAINED WORD EMBEDDING GLOVE DAN  
SUPPORT VECTOR MACHINE DALAM MENDETEKSI BERITA PALSU**



**SKRIPSI**

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

**Brianlian**  
**00000042565**

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

**IMPLEMENTASI PRETRAINED WORD EMBEDDING GLOVE DAN  
SUPPORT VECTOR MACHINE DALAM MENDETEKSI BERITA PALSU**



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

**Brianlian**

**00000042565**

**UMMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

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

**TANGERANG**

**2024**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Brianlian  
Nomor Induk Mahasiswa : 00000042565  
Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Skripsi saya yang berjudul:

### **Implementasi Pretrained Word Embedding GloVe dan Support Vector Machine dalam Mendeteksi Berita Palsu**

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, 10 Mei 2024



(Brianlian)

UMM  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PENGESAHAN

Skripsi dengan judul

### IMPLEMENTASI PRETRAINED WORD EMBEDDING GLOVE DAN SUPPORT VECTOR MACHINE DALAM MENDETEKSI BERITA PALSU

oleh

Nama : Brianlian  
NIM : 00000042565  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Senin, 27 Mei 2024

Pukul 13.00 s/d 15.00 dan dinyatakan

**LULUS**

Dengan susunan penguji sebagai berikut

Ketua Sidang



(Wirawan Istiono, S.Kom., M.Kom.) (Angga Aditya Permiana, S.Kom., M.Kom.)

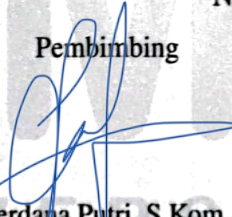
NIDN: 0313048304

Penguji



NIDN: 0407128901

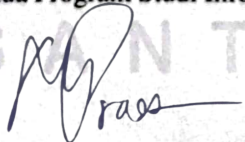
Pembimbing



(Farica Perdana Putri, S.Kom., M.Sc.)

NIDN: 0331019301

Pjs Ketua Program Studi Informatika,



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

NIDN: 0419128203

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Brianlian  
NIM : 00000042565  
Program Studi : Informatika  
Jenjang : S1  
Judul Karya Ilmiah : Implementasi Pretrained Word Embedding  
Glove dan Support Vector Machine Dalam  
Mendeteksi Berita Palsu

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

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

Tangerang, 6 Juni 2024

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

  
(Brianlian)

\* Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

## Halaman Persembahan / Motto

"I have no enemies"

Thorfinn Karlsefni (Vinland Saga)

"Tough times create strong men, strong men create easy times, easy times create weak men, weak men create tough times. You have to raise a warrior"

Aaron McKie

# UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Implementasi Pretrained Word Embedding GloVe dan Support Vector Machine dalam Mendeteksi Berita Palsu 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 Universitas Multimedia Nusantara.
3. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Pjs Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Farica Perdana Putri, S.Kom., M.Sc., sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan, dan motivasi atas terselesainya tesis ini.
5. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.

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

Tangerang, 10 Mei 2024



Brianlian

# IMPLEMENTASI PRETRAINED WORD EMBEDDING GLOVE DAN SUPPORT VECTOR MACHINE DALAM MENDETEKSI BERITA PALSU

Brianlian

## ABSTRAK

Berita palsu merupakan salah satu masalah yang disebabkan karena perkembangan teknologi. Sebagian besar berita palsu yang ada di internet tersebar dalam bentuk teks. Dalam memproses data teks perlu melakukan *feature extraction* untuk mengubah data teks menjadi vektor sehingga dapat diklasifikasikan. Metode vektorisasi kata yang digunakan dalam *feature extraction* teks berita sangat berpengaruh terhadap hasil klasifikasi. Metode vektorisasi kata yang digunakan pada penelitian ini yaitu *term frequency-inverse document frequency* (TF-IDF), Word2Vec, dan GloVe. GloVe adalah model *unsupervised learning* yang memanfaatkan statistik kemunculan kata secara global pada suatu korpus. Algoritma klasifikasi *support vector machine* (SVM) digunakan untuk melakukan klasifikasi berita asli atau palsu terhadap data tersebut. SVM bekerja dengan cara memetakan data ke dalam sebuah ruang vektor yang lalu membentuk sebuah *hyperplane* optimal untuk memisahkan masing-masing kelas. Upaya untuk memastikan bahwa *hyperplane* yang dibentuk oleh SVM dapat memisahkan kelas dengan baik adalah dengan melakukan *hyperparameter tuning*. Proses ini dilakukan setelah *training* dengan cara membandingkan kombinasi kandidat *hyperparameter* sehingga kombinasi tersebut dapat menghasilkan *hyperplane* paling optimal. Performa TF-IDF dengan SVM menempati urutan pertama memiliki akurasi sebesar 97.86%, presisi sebesar 97.39%, *recall* sebesar 98.42%, dan F1-score sebesar 97.92%. Performa GloVe dengan SVM menempati urutan kedua setelah TF-IDF memiliki akurasi sebesar 90.61%, presisi sebesar 89.81%, *recall* sebesar 92.06%, dan F1-score sebesar 90.92%. Word2Vec menempati urutan ketiga memiliki akurasi sebesar 89.17%, presisi sebesar 88.85%, *recall* sebesar 90.08%, dan F1-score sebesar 89.46%.

**Kata kunci:** Berita palsu, GloVe, SVM, TF-IDF, Word2Vec, *word embedding*

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



# Implementation of GloVe Pretrained Word Embedding and Support Vector Machine in Fake News Detection

Brianlian

## ABSTRACT

Fake news is one of the problems caused by technological advancements. Most fake news on the internet is distributed in the form of text. Processing text data requires feature extraction to transform text data into vectors so that it can be classified. The word vectorization method used in news text feature extraction greatly influences the classification results. The word vectorization methods used in this study are Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec, and GloVe. GloVe is an unsupervised learning model that utilizes the statistics of word occurrences globally in a corpus. Support Vector Machine (SVM) classification algorithm is used to classify real or fake news from the data. SVM works by mapping data into a vector space and then forming an optimal hyperplane to separate each class. Ensuring hyperplane formed by SVM can separate the classes well, hyperparameter tuning is required. This process is done after training by comparing candidate hyperparameter combinations so that they can produce the most optimal hyperplane. TF-IDF performance with SVM ranks first with an accuracy of 97.86%, precision of 97.39%, recall of 98.42%, and F1-score of 97.92%. GloVe performance with SVM ranks second after TF-IDF with an accuracy of 90.61%, precision of 89.81%, recall of 92.06%, and F1-score of 90.92%. Word2Vec ranks third with an accuracy of 89.17%, precision of 88.85%, recall of 90.08%, and F1-score of 89.46%.

**Keywords:** Fake news, GloVe, SVM, TF-IDF, Word2Vec, word embedding

U M N  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## 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 GAMBAR . . . . .	xi
DAFTAR TABEL . . . . .	xii
DAFTAR KODE . . . . .	xiii
DAFTAR LAMPIRAN . . . . .	xiv
<b>BAB 1 PENDAHULUAN . . . . .</b>	<b>1</b>
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
<b>BAB 2 LANDASAN TEORI . . . . .</b>	<b>6</b>
2.1 Berita Palsu . . . . .	6
2.2 Word Embedding Algorithm . . . . .	6
2.2.1 Word2Vec . . . . .	7
2.2.2 GloVe . . . . .	9
2.3 Support Vector Machine . . . . .	9
2.4 Confusion Matrix . . . . .	13
<b>BAB 3 METODOLOGI PENELITIAN . . . . .</b>	<b>15</b>
3.1 Metodologi Penelitian . . . . .	15
3.1.1 Gambaran Umum . . . . .	15
3.1.2 Studi Literatur . . . . .	15
3.1.3 Akuisisi Data . . . . .	15
3.1.4 Perancangan Sistem . . . . .	16
3.1.5 Pembangunan Model . . . . .	16
3.1.6 Training, Testing, dan Evaluasi Model . . . . .	17
3.1.7 Penulisan Laporan . . . . .	17
3.2 Perancangan Sistem . . . . .	17
3.2.1 Gambaran Umum . . . . .	17
3.2.2 Preprocessing . . . . .	18
3.2.3 Implement Word Embedding . . . . .	19
3.2.4 Train Test Split . . . . .	20
3.2.5 Implement SVM . . . . .	20
3.2.6 Evaluation . . . . .	21
<b>BAB 4 HASIL DAN DISKUSI . . . . .</b>	<b>22</b>
4.1 Implementasi Sistem . . . . .	22
4.1.1 Exploratory Data Analysis . . . . .	22
4.1.2 Preprocessing . . . . .	25
4.1.3 Implement Word Embedding . . . . .	26

4.1.4	Train Test Split . . . . .	28
4.1.5	Implement SVM . . . . .	28
4.1.6	Evaluation . . . . .	29
4.2	Uji Coba . . . . .	30
4.2.1	Pencarian Hyperparameters Terbaik dengan GridSearchCV . . . . .	30
4.2.2	Perbandingan Performa Metode Vektorisasi Kata . . . . .	31
4.2.3	Perbandingan Performa Metode Meratakan Dimensi GloVe dan Word2Vec . . . . .	32
4.3	Analisis . . . . .	34
BAB 5	SIMPULAN DAN SARAN . . . . .	37
5.1	Simpulan . . . . .	37
5.2	Saran . . . . .	37
	DAFTAR PUSTAKA . . . . .	39



## DAFTAR GAMBAR

Gambar 2.1	Model Skip-gram . . . . .	7
Gambar 2.2	Model CBOW . . . . .	8
Gambar 2.3	SVM Dua Dimensi . . . . .	10
Gambar 2.4	SVM Linear Kernel . . . . .	11
Gambar 2.5	SVM Polynomial Kernel . . . . .	12
Gambar 2.6	SVM RBF Kernel . . . . .	12
Gambar 2.7	Confusion Matrix . . . . .	13
Gambar 3.1	Flowchart Gambaran Umum Metodologi Penelitian . . . . .	15
Gambar 3.2	Flowchart Gambaran Umum Perancangan Sistem . . . . .	17
Gambar 3.3	Flowchart Text Preprocessing . . . . .	18
Gambar 3.4	Flowchart Implementasi Word Embedding . . . . .	20
Gambar 3.5	Flowchart Implementasi SVM . . . . .	21
Gambar 4.1	Pie Chart Kesetimbangan Dataset . . . . .	22
Gambar 4.2	Frekuensi kata berita asli sebelum preprocessing . . . . .	23
Gambar 4.3	Frekuensi kata berita palsu sebelum preprocessing . . . . .	23
Gambar 4.4	Frekuensi kata berita asli setelah preprocessing . . . . .	24
Gambar 4.5	Frekuensi kata berita palsu setelah preprocessing . . . . .	24
Gambar 4.6	Confusion matrix untuk TF-IDF & SVM . . . . .	31
Gambar 4.7	Confusion matrix untuk GloVe & SVM . . . . .	31
Gambar 4.8	Confusion matrix untuk Word2Vec & SVM . . . . .	32



## DAFTAR TABEL

Tabel 2.1	<i>Co-occurrence Matrix</i> . . . . .	9
Tabel 4.1	Tabel kernel & hyperparameters terbaik . . . . .	30
Tabel 4.2	Tabel perbandingan performa metode vektorisasi kata . . . . .	32
Tabel 4.3	Tabel perbandingan performa metode meratakan dimensi word embeddings . . . . .	33
Tabel 4.4	Tabel perbandingan hasil penelitian . . . . .	34
Tabel 4.5	Tabel perbandingan penggunaan hyperparameters SVM . . . . .	35



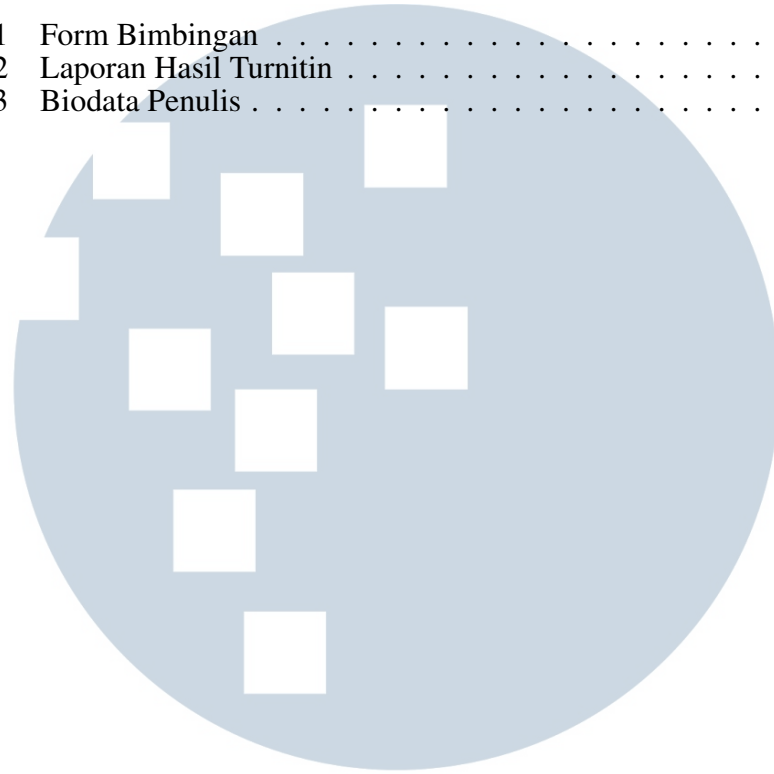
## DAFTAR KODE

Kode 4.1	Potongan kode preprocessing data teks . . . . .	25
Kode 4.2	Potongan kode tokenization & removing stopwords . . . . .	25
Kode 4.3	Potongan kode TF-IDF Vectorizer . . . . .	26
Kode 4.4	Potongan kode GloVe word embedding . . . . .	26
Kode 4.5	Potongan kode Word2Vec word embedding . . . . .	27
Kode 4.6	Potongan kode train test split . . . . .	28
Kode 4.7	Potongan kode SVM . . . . .	28
Kode 4.8	Potongan kode GridSearchCV . . . . .	29
Kode 4.9	Potongan kode confusion matrix . . . . .	29
Kode 4.10	Potongan kode metode meratakan dimensi word embeddings .	33



## DAFTAR LAMPIRAN

Lampiran 1	Form Bimbingan . . . . .	42
Lampiran 2	Laporan Hasil Turnitin . . . . .	43
Lampiran 3	Biodata Penulis . . . . .	47



# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA