

**PERBANDINGAN AKURASI *WORD EMBEDDING WORD2VEC*,  
*GLOVE*, DAN *FASTTEXT* MENGGUNAKAN *SUPPORT VECTOR*  
*MACHINE* UNTUK ANALISIS SENTIMEN ULASAN APLIKASI  
*SPOTIFY***

**SKRIPSI**



**MARGARETHA ANJANI**

**NIM. 1910511108**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**2023**

**PERBANDINGAN AKURASI *WORD EMBEDDING WORD2VEC*,  
*GLOVE*, DAN *FASTTEXT* MENGGUNAKAN *SUPPORT VECTOR*  
*MACHINE* UNTUK ANALISIS SENTIMEN ULASAN APLIKASI  
*SPOTIFY***

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Komputer**



**MARGARETHA ANJANI**

**NIM. 1910511108**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**2023**

## LEMBAR PERSETUJUAN

Yang bertanda tangan di bawah ini:

Nama : Margaretha Anjani

NIM. : 1910511108

Program Studi : Informatika/~~Sistem Informasi~~ Program Sarjana/~~Diploma 3~~ (\*Coret yang tidak perlu)

Judul Skripsi/TA. : Perbandingan Akurasi Word Embedding Word2Vec, GloVe, dan FastText Menggunakan Support Vector Machine Untuk Analisis Sentimen Ulasan Aplikasi Spotify

Dinyatakan telah memenuhi syarat dan menyetujui untuk mengikuti ujian sidang skripsi.


Jakarta, 19 Mei 2023

Mengetahui,  
Ketua Program Studi,



Dr. Widya Cholil, M.I.T.

Menyetujui,  
Dosen Pembimbing,



Helena Nurramdhani Irmanda, S.Pd., M.Kom.

## PERNYATAAN ORISINALITAS

Tugas akhir ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Margaretha Anjani

NIM : 1910511108

Tanggal : 05 Juni 2023

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya, maka saya bersedia di tuntutan dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 05 Juni 2023

Yang menyatakan,



(Margaretha Anjani)

## PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademis Universitas Pembangunan Nasional Veteran Jakarta,  
saya yang bertanda tangan di bawah ini:

Nama : Margaretha Anjani

NIM : 1910511108

Fakultas : Ilmu Komputer

Program Studi : S-1 Informatika

Demi pembangunan ilmu pengetahuan, menyetujui untuk memberikan pelayanan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas karya ilmiah saya yang berjudul :

**Perbandingan Akurasi Word Embedding Word2Vec, GloVe, Dan FastText  
Menggunakan Support Vector Machine Untuk Analisis Sentimen Ulasan  
Aplikasi Spotify**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tugas Akhir saya selama mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat : Jakarta

Pada tanggal : 05 Juni 2023

Yang menyatakan



(Margaretha Anjani)

## LEMBAR PENGESAHAN

Yang bertanda tangan di bawah ini:

Nama : Margaretha Anjani

NIM : 1910511108

Program Studi : S-1 Informatika

Judul Skripsi/TA : Perbandingan Akurasi Word Embedding Word2Vec, GloVe, Dan FastText Menggunakan Support Vector Machine Untuk Analisis Sentimen Ulasan Aplikasi Spotify

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



(Dr. Bambang Saras, S.T., M.Kom)

Penguji 1



(Ati Zaidiah, S.Kom., MTL.)

Penguji 2



(Helena Nurramdhani Irmanda,

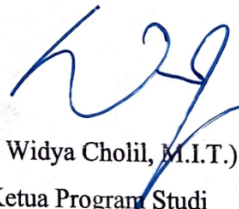
S.Pd., M.Kom.)

Pembimbing



(Dr. Ermatita, M.Kom.)

Dekan



(Dr. Widya Cholil, M.I.T.)

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 26 Mei 2023



Perbandingan Akurasi *Word Embedding Word2Vec, GloVe, dan FastText* Menggunakan  
*Support Vector Machine* untuk Analisis Sentimen Ulasan Aplikasi Spotify

Margaretha Anjani

Program Studi Informatika, Fakultas Ilmu Komputer,  
Universitas Pembangunan Nasional Veteran Jakarta

Email : [margarethanjani@upnvj.ac.id](mailto:margarethanjani@upnvj.ac.id)

## ABSTRAK

Spotify merupakan salah satu aplikasi yang digunakan sebagai platform layanan streaming audio digital yang menyajikan berbagai musik dan podcast dan dapat diunduh dengan gratis pada Google Play Store. Ulasan merupakan sebuah fitur Google Play Store yang dapat dimanfaatkan oleh pengguna untuk memberikan penilaian terhadap sebuah aplikasi. Ulasan yang dapat diterima oleh aplikasi dapat mempengaruhi pengguna yang akan mengunduh aplikasi tersebut. Karakteristik teks ulasan yang tidak terstruktur akan menjadi sebuah tantangan dalam proses pemrosesan teks. Untuk menghasilkan sentimen analisis yang valid dibutuhkan adanya penerapan *word embedding*. Data set yang dimiliki dibagi dengan perbandingan 80:20 untuk data training dan data testing. Metode yang digunakan untuk ekspansi fitur *Word2Vec, GloVe, dan FastText* dan metode yang digunakan dalam klasifikasi adalah *Support Vector Machine (SVM)*. Ketiga metode *word embedding* tersebut dipilih karena dapat menangkap makna yang semantik, sintatik, serta konteks pada sekitar kata bila dibandingkan *feature engineering* tradisional seperti *Bag of Word*. Hasil evaluasi performa terbaik menunjukkan model GloVe menghasilkan kinerja terbaik dibandingkan dengan word embedding lainnya dengan nilai akurasi sebesar 85%, nilai presisi sebesar 90%, nilai *recall* 79%, dan *f1-score* 85%.

**Kata kunci:** word2vec, glove, fasttext, support vector machine, klasifikasi, analisis sentimen.

*Comparison Accuracy of Word Embedding Word2Vec, GloVe, dan FastText Using Support Vector Machine for Sentiment Analysis Spotify App Reviews*

Margaretha Anjani

Program Studi Informatika, Fakultas Ilmu Komputer,  
Universitas Pembangunan Nasional Veteran Jakarta

Email : [margarethanjani@upnvj.ac.id](mailto:margarethanjani@upnvj.ac.id)

**ABSTRACT**

*Spotify is an application that is used as a digital audio streaming service platform that presents a variety of music and podcasts and can be downloaded for free on the Google Play Store. Reviews are a feature of the Google Play Store that can be used by users to rate an application. The reviews that an app can receive may affect the users who will download the app. Characteristics of unstructured review texts will be a challenge in the text processing process. To produce a valid sentiment analysis, it is necessary to apply word embedding. The data set that is owned is divided by a ratio of 80:20 for training data and testing data. The method used for the expansion of the Word2Vec, GloVe, and FastText features and the method used in the classification is the Support Vector Machine (SVM). The three word embedding methods were chosen because they can capture the semantic, syntactic, and contextual meanings around words when compared to traditional engineering features such as Bag of Word. The best performance evaluation results show that the GloVe model produces the best performance compared to other word embeddings with an accuracy value of 85%, a precision value of 90%, a recall value of 79%, and an f1-score of 85%.*

**Keywords:** *word2vec, glove, fasttext, support vector machine, classification, sentiment analysis.*



## KATA PENGANTAR

Dengan mengucapkan syukur dan puji kehadiran Tuhan Yang Maha Esa yang telah memberikan rahmat, kesehatan, petunjuk, dan pertolongan, serta semua karunia-Nya, sehingga peneliti dapat menyelesaikan skripsi ini yang berjudul “Perbandingan Akurasi *Word Embedding Word2Vec, GloVe, FastText* Menggunakan *Support Vector Machine* Untuk Analisis Sentimen Ulasan Aplikasi Spotify”.

Skripsi ini disusun peneliti guna memenuhi salah satu syarat untuk memperoleh Gelar Sarjana pada Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta. Pada penyusunan skripsi ini, peneliti memperoleh banyak dukungan serta bantuan dalam berbagai aspek baik spiritual, moral, dan material yang telah didapatkan peneliti selama proses skripsi sehingga akhirnya mampu menyelesaikan penelitian skripsi ini dengan baik dan lancar. Dengan rasa hormat, peneliti mengucapkan terima kasih kepada :

1. Ibu, Bapak, Kakak, dan seluruh keluarga peneliti yang telah senantiasa memberikan doa, dukungan, dan semangat sehingga peneliti dapat menyelesaikan penelitian dan skripsi.
2. Ibu Helena Nurramdhani Irmanda, S.Pd., M.Kom. sebagai dosen pembimbing yang telah senantiasa memberikan semangat dan membimbing serta memberi arahan untuk peneliti dalam proses menyusun skripsi.
3. Ibu Dr. Widya Cholil, M.I.T. sebagai Ketua Program Studi Informatika.
4. Ibu Dr. Ermatita, M.Kom. sebagai Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
5. Bapak/Ibu Dosen Fakultas Ilmu Komputer yang berdedikasi mengajar dan membagikan ilmu sehingga peneliti dapat memperluas wawasan serta ilmu pengetahuan selama berstatus menjadi mahasiswa di kawasan Fakultas Ilmu Komputer.
6. Sahabat-sahat peneliti yang telah senantiasa mendukung, mengajak, dan mendoakan peneliti dalam melakukan penelitian dan penyusunan skripsi.
7. Teman-teman peneliti dan semua pihak lain yang tidak dapat disebutkan satu persatu atas semua kontribusi untuk peneliti dalam penelitian dan penyusunan skripsi baik secara langsung maupun tidak langsung.

Peneliti sangat menyadari masih terdapat banyak kekurangan dan kesalahan pada skripsi ini mengingat dengan keterbatasan pengetahuan dan kemampuan yang dimiliki oleh peneliti. Oleh karena itu, peneliti sangat mengharapkan skripsi ini dapat memberikan informasi dan pengetahuan yang berharga untuk pembaca.

Jakarta, 10 Mei 2023

Peneliti,

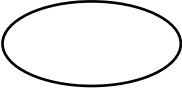
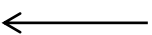

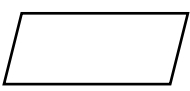
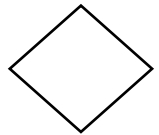
# DAFTAR ISI

LEMBAR PERSETUJUAN .....	iii
PERNYATAAN ORISINALITAS .....	iv
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS .....	v
LEMBAR PENGESAHAN .....	vi
ABSTRAK.....	vii
<i>ABSTRACT</i> .....	viii
KATA PENGANTAR .....	ix
DAFTAR ISI.....	xi
DAFTAR SIMBOL .....	xiv
DAFTAR GAMBAR.....	xv
DAFTAR TABEL.....	xvi
DAFTAR LAMPIRAN.....	xvii
BAB I.....	1
PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	4
1.3 Ruang Lingkup Penelitian.....	5
1.4 Tujuan Penelitian .....	5
1.5 Manfaat Penelitian .....	6
1.6 Luaran yang Diharapkan.....	6
1.7 Sistematika Penelitian.....	6
BAB II.....	9
TINJAUAN PUSTAKA .....	9
2.1 Analisis Sentimen .....	9
2.2 Spotify.....	9
2.3 <i>Web Scraping</i> .....	10
2.4 <i>Text Mining</i> .....	10
2.5 <i>Text Preprocessing</i> .....	11
2.5.1 <i>Case Folding</i> .....	11
2.5.2 <i>Data Cleaning</i> .....	11
2.5.3 <i>Tokenizing</i> .....	11

2.5.4 <i>Normalization</i> .....	12
2.5.5 <i>Stopword Removal</i> .....	12
2.5.6 <i>Stemming</i> .....	12
2.6 <i>Word Embedding</i> .....	12
2.6.1 <i>Word2Vec</i> .....	13
2.6.2 <i>GloVe</i> .....	14
2.6.3 <i>FastText</i> .....	15
2.7 <i>Support Vector Machine</i> .....	16
2.7.1 <i>SVM Linear</i> .....	17
2.7.2 <i>SVM Non-Linear</i> .....	18
2.8 <i>Confusion Matrix</i> .....	19
2.9 <i>Penelitian Terdahulu</i> .....	21
<b>BAB III</b> .....	<b>24</b>
<b>METODE PENELITIAN</b> .....	<b>24</b>
3.1 <i>Analisa dan Desain Tahapan Penelitian</i> .....	24
3.2 <i>Menentukan Topik</i> .....	25
3.3 <i>Identifikasi Masalah</i> .....	26
3.4 <i>Studi Pustaka</i> .....	26
3.5 <i>Pengumpulan Data dan Labelling</i> .....	26
3.6 <i>Preprocessing</i> .....	27
3.6.1 <i>Case Folding</i> .....	28
3.6.2 <i>Data Cleaning</i> .....	29
3.6.3 <i>Tokenizing</i> .....	30
3.6.4 <i>Normalization</i> .....	32
3.6.5 <i>Stopword Removal</i> .....	33
3.6.6 <i>Stemming</i> .....	34
3.7 <i>Ekstraksi Fitur</i> .....	35
3.8 <i>Klasifikasi Data</i> .....	36
3.9 <i>Pengujian Sistem</i> .....	37
3.10 <i>Perangkat yang Digunakan</i> .....	37
3.11 <i>Jadwal Penelitian</i> .....	38
<b>BAB IV</b> .....	<b>39</b>
<b>PEMBAHASAN</b> .....	<b>39</b>
4.1 <i>Pengumpulan Data dan Labelling</i> .....	39
4.2 <i>Preprocessing</i> .....	41
4.2.1 <i>Case Folding</i> .....	42

4.2.2	<i>Data Cleaning</i>	43
4.2.3	<i>Tokenizing</i>	45
4.2.4	<i>Normalization</i>	46
4.2.5	<i>Stopword Removal</i>	48
4.2.6	<i>Stemming</i>	49
4.3	Pembagian Data	50
4.4	Ekstraksi Fitur	51
4.4.1	<i>Word2Vec</i>	52
4.4.2	<i>GloVe</i>	53
4.4.3	<i>FastText</i>	54
4.5	Klasifikasi	56
4.6	Evaluasi Klasifikasi Model	56
4.6.1	Evaluasi Klasifikasi Model <i>Word2Vec</i>	57
4.6.2	Evaluasi Klasifikasi Model <i>GloVe</i>	60
4.6.3	Evaluasi Klasifikasi Model <i>FastText</i>	62
4.6.4	Perbandingan Performa Klasifikasi Model Word Embedding	65
BAB V		70
PENUTUP		70
5.1	Kesimpulan	70
5.2	Saran	71
DAFTAR PUSTAKA		73
RIWAYAT HIDUP		76
LAMPIRAN		77

## DAFTAR SIMBOL

No	Gambar	Nama	Keterangan
1		<i>Terminal (start, end)</i>	Menampilkan sebuah kegiatan dimulai atau kegiatan berakhir
2		<i>Flow direction</i>	Menampilkan alur dan arah sebuah proses
3		<i>Process</i>	Menampilkan sebuah proses yang dilakukan
4		<i>Data</i>	Menampilkan sebuah proses yang dilakukan memiliki data masukan ( <i>input</i> ) atau data keluaran ( <i>output</i> )
5		<i>Decision</i>	Menampilkan sebuah proses yang memiliki kondisi yang harus memilih lebih dari satu proses

## DAFTAR GAMBAR

Gambar 2.1 CBOW.....	13
Gambar 2.2 Skip-gram.....	14
Gambar 2.3 Arsitektur Model FastText.....	16
Gambar 2.4 Pemisah Hyperplane.....	17
Gambar 3.1 Tahapan Penelitian.....	24
Gambar 3.2 Pengumpulan Data & Labelling.....	26
Gambar 3.3 Tahapan Preprocessing.....	27
Gambar 3.4 Case Folding.....	28
Gambar 3.5 Data Cleaning.....	29
Gambar 3.6 Tokenizing.....	31
Gambar 3.7 Normalization.....	32
Gambar 3.8 Stopword Removal.....	33
Gambar 3.9 Stemming.....	35
Gambar 3.10 Tahapan Klasifikasi.....	36
Gambar 4.1 Hasil Scraping Data Ulasan Spotify.....	39
Gambar 4.2 Grafik Pelabelan Data Ulasan.....	41
Gambar 4.3 Hasil Case Folding.....	43
Gambar 4.4 Hasil Data Cleaning.....	45
Gambar 4.5 Hasil Tokenizing.....	46
Gambar 4.6 Hasil Normalization.....	47
Gambar 4.7 Hasil Stopword Removal.....	49
Gambar 4.8 Hasil Stemming.....	50
Gambar 4.9 Perbandingan Hasil Evaluasi Model dengan Data Testing.....	68
Gambar 4.10 Perbandingan Hasil Evaluasi Model dengan Data Training.....	69

## DAFTAR TABEL

Tabel 2.1 Confusion Matrix .....	19
Tabel 2.2 Penelitian Terdahulu .....	21
Tabel 3.1 Perubahan Sebelum dan Sesudah Case Folding .....	28
Tabel 3.2 Perubahan Sebelum dan Sesudah Data Cleaning .....	30
Tabel 3.3 Perubahan Sebelum dan Sesudah Tokenizing .....	31
Tabel 3.4 Perubahan Sebelum dan Sesudah Normalization .....	32
Tabel 3.5 Perubahan Sebelum dan Sesudah Stopword Removal .....	34
Tabel 3.6 Perubahan Sebelum dan Sesudah Stemming .....	35
Tabel 3.7 Jadwal Penelitian .....	38
Tabel 4.1 Hasil Pelabelan Pada Sampel Data Ulasan .....	40
Tabel 4.2 Jumlah Pelabelan Pada Seluruh Data Ulasan .....	40
Tabel 4.3 Sebelum dan Sesudah Case Folding .....	42
Tabel 4.4 Sebelum dan Sesudah Data Cleaning .....	43
Tabel 4.5 Sebelum dan Sesudah Tokenizing .....	45
Tabel 4.6 Sebelum dan Sesudah Normalization .....	46
Tabel 4.7 Sebelum dan Sesudah Stopword Removal .....	48
Tabel 4.8 Sebelum dan Sesudah Stemming .....	49
Tabel 4.9 Pembagian Data Testing dan Data Training .....	51
Tabel 4.10 Nilai Vector Hasil Model Word2Vec .....	53
Tabel 4.11 Nilai Vector Hasil Model GloVe .....	54
Tabel 4.12 Nilai Vector Hasil Model FastText.....	55
Tabel 4.13 Hasil Nilai Akurasi .....	56
Tabel 4.14 Confusion Matrix Klasifikasi Model Word2Vec dengan Data Testing .....	57
Tabel 4.15 Confusion Matrix Klasifikasi Model Word2Vec dengan Data Training.....	58
Tabel 4.16 Confusion Matrix Klasifikasi Model GloVe dengan Data Testing .....	60
Tabel 4.17 Confusion Matrix Klasifikasi Model GloVe dengan Data Training.....	61
Tabel 4.18 Confusion Matrix Klasifikasi Model FastText dengan Data Testing .....	63
Tabel 4.19 Confusion Matrix Klasifikasi Model FastText dengan Data Training .....	64
Tabel 4.20 Perbandingan Performa Model Word Embedding.....	66



## DAFTAR LAMPIRAN

Lampiran 1. Kamus Normalization.....	77
Lampiran 2. Daftar Stopword .....	85
Lampiran 3. Source Code .....	87
Lampiran 4. Uji Turnitin.....	94