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,	Menyetujui,
Ketua Program Studi,	Dosen Pembimbing,
Dr. Widya Cholil, M.I.T.	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 ketidaksesuain dengan pernyataan saya, maka saya bersedia di tuntut 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-ekslusif (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 Royalty 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., MTI.)
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

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

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 kehadirat 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 menyususn 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.

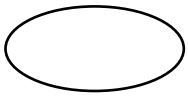
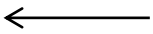
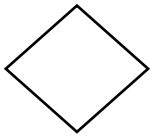
Peneliti sangat menyadari masih terdapat banyak kekurangan dan kesalahan pada skripsi ini mengingat dengan keterbatasan pengetahuan dan kemampuan yang dimiliki oleh peneliti. Oleh karena itu, peneliti sangat mengharapkan skripsi ini dapat memberikan informasi dan pengetahuan yang berharga untuk pembaca.

Jakarta, 10 Mei 2023

Peneliti,

DAFTAR ISI

DAFTAR SIMBOL

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

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN