

KATA PENGANTAR

Segala puji dan syukur kepada Allah SWT, karena berkat rahmat, petunjuk, dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul “Normalisasi *Word Vector* Pada *Word Embedding*: Pendekatan Aljabar untuk Meningkatkan Kualitas Representasi Kata”. Skripsi ini disusun untuk memenuhi salah satu syarat ujian sarjana pada Program Studi S-1 Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Padjadjaran.

Skripsi ini tidak dapat selesai tanpa adanya dukungan dan bantuan dari berbagai pihak. Untuk itu pada kesempatan ini, penulis mengucapkan terima kasih kepada Dr. Sisilia Sylviani, S.Si., M.Si., selaku Dosen Pembimbing Utama dan Dr. Anita Triska, S.Si., M.Si., selaku Dosen Pembimbing Pendamping yang telah memberikan bimbingan, bantuan, dan dorongan yang sangat berharga kepada penulis dalam proses penyusunan skripsi ini. Selain itu, penulis juga ingin mengucapkan terima kasih kepada:

1. Prof. Dr. Iman Rahayu, S.Si., M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Padjadjaran.
2. Dr. Ema Carnia, M.Si., selaku Kepala Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Padjadjaran.
3. Edi Kurniadi, S.Si., M.Si., Ph.D., selaku Ketua Program Studi S-1 Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Padjadjaran
4. Prof. Dr. Asep Kuswandi Supriatna, Ms., selaku Dosen Wali.

5. Seluruh Civitas Akademika Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Padjadjaran.
6. Kedua orang tua, kakak serta keluarga besar penulis yang selalu memberikan dukungan, motivasi serta doa yang tidak pernah terputus kepada penulis.
7. Seluruh pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan banyak dukungan dan doa kepada penulis.

Semoga skripsi ini dapat bermanfaat bagi diri penulis sendiri, orang-orang yang membacanya, dan bahkan masyarakat secara luas lewat segala ilmu dan gagasannya.

Bandung, Agustus 2023

Penulis

DAFTAR ISI

ABSTRAK	iii
ABSTRACT	iv
KATA PENGANTAR	v
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR NOTASI	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	5
1.5 Kegunaan Penelitian	6
1.6 Metodologi Penelitian	6
1.7 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Matriks	9
2.2 Ruang Vektor	10
2.3 Hasil kali Dalam	15
2.4 Ruang Benorma	19
2.5 Ruang Metrik	20
2.6 <i>Word Vector</i>	21
2.7 <i>Word Embedding</i>	24
2.8 <i>Word2Vec</i>	26
2.9 <i>One-Hot Encoding</i>	29
2.10 Fungsi Softmax	30
2.11 Nilai <i>Loss</i>	32
2.12 <i>Hypersphere</i>	32
BAB III OBJEK DAN METODE PENELITIAN	34

3.1	Objek Penelitian.....	34
3.2	Metode Penelitian	34
BAB IV HASIL DAN PEMBAHASAN		37
4.1	Skip-Gram.....	45
4.1.1	Input Layer	47
4.1.2	Hidden Layer.....	47
4.1.3	Output Layer	48
4.1.4	Galat	49
4.2	Normalisasi <i>Word Vector</i>	49
4.3	Studi Kasus Untuk Model Skip-Gram	50
4.4	Studi Kasus Untuk Normalisasi <i>Word Vector</i>	54
4.5	Mengukur Jarak Antara <i>Word Vector</i>	56
BAB V SIMPULAN DAN SARAN		63
5.1	Simpulan	63
5.2	Saran	63
DAFTAR PUSTAKA		65

DAFTAR GAMBAR

Gambar 2.1 Representasi word vector (Bujokas, 2020)	25
Gambar 2.2 Model CBOW (Rong, 2016)	27
Gambar 2.3 Model Skip-gram (Rong, 2016)	28
Gambar 3.1 Diagram Alir Penelitian	35
Gambar 3.2 Diagram Proses <i>Word Embedding</i> dengan model Skip-gram.....	36

DAFTAR TABEL

Tabel 2.1 Contoh Representasi kata pada dokumen V	23
Tabel 4.1 Representasi kata dalam bentuk <i>one-hot encoding</i> vector	51
Tabel 4.2 Jarak antara word vector	62

DAFTAR NOTASI

W	: <i>Weight Matrix</i> antara <i>input layer</i> dan <i>hidden layer</i>
M	: <i>Weight Matrix</i> antara <i>hidden layer</i> dan <i>output layer</i>
v	: Banyaknya kata yang berbeda dalam suatu kalimat/dokumen.
s	: <i>Size</i>
k	: Banyaknya <i>context word</i>
\mathbf{x}_i	: <i>One-hot encoding vector</i> untuk <i>focus word</i>
\mathbf{x}_c	: <i>One-hot encoding vector</i> untuk <i>context word</i>
$x(i, j)$: Entri ke- j pada <i>one-hot encoding vector</i> ke- i
\mathbf{y}	: Vektor perkiraan yang diperoleh dari fungsi softmax
\mathbf{e}_k	: Galat dari <i>context word</i> ke- v
M_c^T	: Baris ke- c dari matriks M^T
\mathbf{h}	: Representasi <i>word vector</i> dari <i>focus word</i> pada matriks W
\mathbf{w}_c	: Representasi <i>word vector</i> dari <i>context word</i> ke- c pada matriks M
\mathbf{w}_i	: Representasi <i>word vector</i> untuk kata ke- i
$w(i, j)$: Entri ke- j pada <i>word vector</i> untuk kata ke- i
i	= 1,2,3, ..., v
c	= 1,2,3, ..., v .