UNIVERSITAS
GADJAH MADA
PERBANDINGAN KOMBINASI WORD EMBEDDING PADA ARSITEKTUR 2D DAN 3D CNN UNTUK KLASIFIKASI TEKS BAHASA INDONESIA
Fitra Ahya Mubarok, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR ISI

HALAMAN PENGESAHAN i
PRAKATA ii
DAFTAR ISI iv
DAFTAR GAMBAR vii
DAFTAR TABEL x
INTISARI xii
ABSTRACT xiii
BAB I PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 4
1.3 Batasan Masalah 4
1.4 Tujuan 5
1.5 Manfaat 5
BAB II TINJAUAN PUSTAKA 7
BAB III METODOLOGI PENELITIAN 18
3.1 Deskripsi Umum Penelitian 18
3.2 Akuisisi Data 18
3.3 Rancangan Metode 26
3.4 Rancangan Pengujian dan Evaluasi 37
BAB IV IMPLEMENTASI 46
4.1 Spesifikasi Perangkat Lunak Dan Keras 46
4.2 Pengumpulan Data 46
4.3 Text Preprocessing 47
4.3.1 Case Folding 48
4.3.2 *Cleaning Noise* 48
4.3.3 *Normalization* 49
4.3.4 *Tokenization* 50

UNIVERSITAS
GADJAH MADA
PERBANDINGAN KOMBINASI WORD EMBEDDING PADA ARSITEKTUR 2D DAN 3D CNN UNTUK KLASIFIKASI TEKS BAHASA INDONESIA
Fitra Ahya Mubarok, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

4.3.5 *Stopword removal* 51
4.3.6 *Stemming* 51
4.3.7 *Padding* 52
4.4 Word Embedding 52
4.4.1 Word2Vec 52
4.4.2 Glove 53
4.4.3 FastText 55
4.5 Training setup 56
4.6 Arsitektur Model 56
4.6.1 Arsitektur R1 (mengacu pada Faisal et al) 56
4.6.2 Arsitektur R2 (Su et al) 58
4.6.3 Sequence *Model* 64
4.7 Evaluasi 65
BAB V HASIL DAN PEMBAHASAN 67
5.1 Hasil Preprocessing 67
5.1.1 EmoT 67
5.1.2 SmSA 67
5.1.3 WReTE 68
5.1.4 *KEPS* 68
5.1.5 TermA 69
5.1.6 FacQA 70
5.2 Word Embedding 71
5.3 Evaluasi Model 73
5.3.1 Hasil Penelitian Setiap Dataset 73
5.3.2 Benchmark Terhadap IndoNLU 133
BAB VI KESIMPULAN DAN SARAN 143
6.1 Kesimpulan 143
6.2 Saran 144

UNIVERSITAS
GADJAH MADA
PERBANDINGAN KOMBINASI WORD EMBEDDING PADA ARSITEKTUR 2D DAN 3D CNN UNTUK KLASIFIKASI TEKS BAHASA INDONESIA
Fitra Ahya Mubarok, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA 145
LAMPIRAN 1. SURAT PERNYATAAN PLAGIARISME..... 148
LAMPIRAN 2. SURAT PERNYATAAN PENGGUNAAN TOOLS AI.... 149
LAMPIRAN 3. DRAFT PAPER/SUBMITTED PAPER..... 150
LAMPIRAN 4. HASIL SIMILARITY PAPER..... 151
LAMPIRAN 5. BUKTI SUBMIT PAPER 152

UNIVERSITAS
GADJAH MADA
PERBANDINGAN KOMBINASI WORD EMBEDDING PADA ARSITEKTUR 2D DAN 3D CNN UNTUK KLASIFIKASI TEKS BAHASA INDONESIA
Fitra Ahya Mubarok, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR GAMBAR

Gambar 3. 1 Diagram tahapan penelitian 26
Gambar 3. 2 Diagram pembuatan model 28
Gambar 3. 3 Diagram alur preprocessing 29
Gambar 3.4 Model Arsitektur 3D CNN (Wang et al., 2023) 33
Gambar 3. 5 Arsitektur CNN 34
Gambar 3. 6 Arsitektur attention 35
Gambar 3. 7 Hasil *Concatenate word embedding* 36
Gambar 3. 8 Arsitektur *Faisal et al.,* 2024 36
Gambar 3.9 Diagram Alur Skenario Pertama 38
Gambar 3.10 Diagram Alur Skenario Kedua 39
Gambar 3.11 Diagram Alur Skenario Ketiga 40
Gambar 3.12 Diagram Alur Skenario Keempat 41
Gambar 3.13 Diagram Alur Skenario Kelima 41
Gambar 3.14 Diagram Alur Skenario Keenam 42
Gambar 3. 15 Diagram Alur Skenario Ketujuh 43
Gambar 3. 16 Diagram Alur Skenario Kedelapan 43
Gambar 3. 17 Diagram Alur Skenario Kesembilan 44
Gambar 3. 18 Diagram Alur Skenario Kesepuluh 45
Gambar 4. 1 Implementasi *case folding* 48
Gambar 4. 2 Implementasi *cleaning noise* 49
Gambar 4. 3 Implementasi *normalization* 50
Gambar 5. 4 Implementasi *tokenization* 51
Gambar 4. 5 Implementasi *stopword* 51
Gambar 4. 6 *Implementasi stemming* 51
Gambar 4. 7 *Implementasi Padding* 52
Gambar 4. 8 *Implementasi Word2Vec* 52
Gambar 4. 9 Implementasi *Glove* 54
Gambar 4. 10 *Implementasi FastText* 55

PERBANDINGAN KOMBINASI WORD EMBEDDING PADA ARSITEKTUR 2D DAN 3D CNN UNTUK KLASIFIKASI TEKS BAHASA INDONESIA
Fitra Ahya Mubarok, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Gambar 4. 11 *Implementasi 2D CNN* Arsitektur Faisal et al (2024) 57
Gambar 4. 12 *Implementasi 3D CNN* Arsitektur Faisal et al (2024) 58
Gambar 4. 13 *Implementasi 2D CNN* Arsitektur Su et al (2024)..... 61
Gambar 4. 14 *Implementasi Attention 2D CNN* Arsitektur Su et al (2024).... 61
Gambar 4. 15 *Implementasi 3D CNN* Arsitektur Su et al (2024)..... 63
Gambar 4. 16 *Implementasi Attention 3D CNN* Arsitektur Su et al (2024).... 64
Gambar 4. 17 *Implementasi Sequence Labelling* pada bagian *LSTM* dan *ConvLSTM* 65
Gambar 4. 18 Implementasi Evaluasi 66
Gambar 5. 1 Hasil training FacQA 79
Gambar 5. 2 *Confusion Matrix* (a) S2 + Word2Vec, (b) S4 + FastText, (c) S4 + Word2Vec, (d) S4 + GloVe, dan (e) S8 85
Gambar 5. 3 Confusion Matrix S6 92
Gambar 5. 4 *Confusion Matrix* (a) S2 + glove, (b) S4 + FastText, (c) S4 + Word2Vec, (d) S4 + GloVe, dan (e) S8 95
Gambar 5. 5 Confusion Matrix S6 98
Gambar 5. 6 *Confusion Matrix* (a) S2 + FastText, (b) S6, (c) S4 + FastText, (d) S4 + Word2Vec, (e) S4 + GloVe, dan (f) S8 109
Gambar 5. 7 Perbandingan total waktu pelatihan 3D pada dataset klasifikasi IndoNLU 117
Gambar 5. 8 Perbandingan total waktu pelatihan 3D pada dataset sequence labeling IndoNLU 118
Gambar 5. 9 Grafik perbandingan total waktu pelatihan 3D pada dataset bencana untuk S2, S4, S6, dan S8 119
Gambar 5. 10 Grafik validation loss per epoch pada dataset EmoT untuk perbandingan S2, S4, S6, dan S8 120
Gambar 5. 11 Grafik akurasi training per epoch pada dataset EmoT untuk perbandingan S2, S4, S6, dan S8 121
Gambar 5. 12 Grafik validation loss per epoch pada dataset SmSA untuk perbandingan S2, S4, S6, dan S8 122

PERBANDINGAN KOMBINASI WORD EMBEDDING PADA ARSITEKTUR 2D DAN 3D CNN UNTUK KLASIFIKASI TEKS BAHASA INDONESIA
Fitra Ahya Mubarok, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Gambar 5. 13 Grafik akurasi training per epoch pada dataset SmSA untuk perbandingan S2, S4, S6, dan S8 122
Gambar 5. 14 Grafik validation loss per epoch pada dataset WReTE untuk perbandingan S2, S4, S6, dan S8 123
Gambar 5. 15 Grafik akurasi training per epoch pada dataset WReTE untuk perbandingan S2, S4, S6, dan S8 124
Gambar 5. 16 Grafik validation loss per epoch pada dataset KEPS untuk perbandingan S2, S4, S6, dan S8 125
Gambar 5. 17 Grafik akurasi training per epoch pada dataset KEPS untuk perbandingan S2, S4, S6, dan S8 125
Gambar 5. 18 Grafik validation loss per epoch pada dataset TermA untuk perbandingan S2, S4, S6, dan S8 126
Gambar 5. 19 Grafik akurasi training per epoch pada dataset TermA untuk perbandingan S2, S4, S6, dan S8 127
Gambar 5. 20 Grafik validation loss per epoch pada dataset FacQA untuk perbandingan S2, S4, S6, dan S8 128
Gambar 5. 21 Grafik akurasi training per epoch pada dataset FacQA untuk perbandingan S2, S4, S6, dan S8 128
Gambar 5. 22 Grafik validation loss per epoch pada dataset Banjir untuk perbandingan S2, S4, S6, dan S8 129
Gambar 5. 23 Grafik akurasi training per epoch pada dataset Banjir untuk perbandingan S2, S4, S6, dan S8 129
Gambar 5. 24 Grafik validation loss per epoch pada dataset Gempa untuk perbandingan S2, S4, S6, dan S8 130
Gambar 5. 25 Grafik akurasi training per epoch pada dataset Gempa untuk perbandingan S2, S4, S6, dan S8 131
Gambar 5. 26 Grafik validation loss per epoch pada dataset Kebakaran untuk perbandingan S2, S4, S6, dan S8 132
Gambar 5. 27 Grafik akurasi training per epoch pada dataset Kebakaran untuk perbandingan S2, S4, S6, dan S8 132

UNIVERSITAS
GADJAH MADA
PERBANDINGAN KOMBINASI WORD EMBEDDING PADA ARSITEKTUR 2D DAN 3D CNN UNTUK KLASIFIKASI TEKS BAHASA INDONESIA
Fitra Ahya Mubarok, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR TABEL

Tabel 2.1 Tinjauan pustaka 11
Tabel 3. 1 Rincian dataset 19
Tabel 3. 2 Contoh data TermA 22
Tabel 3. 3 Contoh data KEPS 23
Tabel 3. 4 Contoh data FacQA 24
Tabel 3. 5 Struktur vector pada word embedding 31
Tabel 3. 6 Struktur vector kalimat 31
Tabel 4. 1 Spesifikasi Perangkat Keras (*hardware*) 46
Tabel 4. 2 Spesifikasi Perangkat Lunak (*software*) 46
Tabel 4. 3 Embedding Training Setup 56
Tabel 4. 4 Model Training Setup 56
Tabel 5. 1 Pembagian label pada dataset EmoT 67
Tabel 5. 2 Statistik dataset EmoT 67
Tabel 5. 3 Pembagian label pada dataset SmSA 68
Tabel 5. 4 Statistik dataset SmSA 68
Tabel 5. 5 Pembagian label pada dataset WReTE 68
Tabel 5. 6 Statistik dataset WReTE 68
Tabel 5. 7 Pembagian label token pada dataset KEPS 69
Tabel 5. 8 Pembagian label entitas pada dataset KEPS 69
Tabel 5. 9 Statistik dataset KEPS 69
Tabel 5. 10 Pembagian label pada dataset TermA 69
Tabel 5. 11 Pembagian label entitas pada dataset TermA 69
Tabel 5. 12 Statistik dataset TermA 70
Tabel 5. 13 Pembagian label pada dataset FacQA 70
Tabel 5. 14 Pembagian label entitas pada dataset FacQA 70
Tabel 5. 15 Statistik dataset FacQA 70
Tabel 5. 16 *Sanity Check* pada data TermA 71
Tabel 5. 17 *Sanity Check* pada data EmoT 72

UNIVERSITAS
GADJAH MADA
PERBANDINGAN KOMBINASI WORD EMBEDDING PADA ARSITEKTUR 2D DAN 3D CNN UNTUK KLASIFIKASI TEKS BAHASA INDONESIA
Fitra Ahya Mubarok, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Tabel 5. 18 Hasil Setiap Dataset *Single Task Classification* 74
Tabel 5. 19 Hasil Setiap Dataset *Sequence Labeling* 75
Tabel 5. 20 Hasil Distribusi Kelas Dataset EmoT 86
Tabel 5. 21 Jumlah Data Mengandung Kata "*takut*" 88
Tabel 5. 22 Distribusi hasil prediksi 89
Tabel 5. 23 Rata-rata probabilitas prediksi 89
Tabel 5. 24 Contoh Data Mengandung Kata "*takut*" 89
Tabel 5. 25 Hasil Distribusi Kelas Dataset SmSA 95
Tabel 5. 26 Distribusi contoh kata pada SmSA 99
Tabel 5. 27 Contoh kalimat positive pada SmSA 100
Tabel 5. 28 Contoh kalimat neutral pada SmSA 103
Tabel 5. 29 Contoh kalimat negative pada SmSA 104
Tabel 5. 30 Contoh kalimat negative pada SmSA Lanjutan 105
Tabel 5. 31 Hasil Distribusi Kelas Dataset Banjir 109
Tabel 5. 32 Contoh data don't know 110
Tabel 5. 33 Contoh data noneyewitness 111
Tabel 5. 34 Contoh data eyewitness 113
Tabel 5. 35 Hasil Benchmark Pada Setiap Dataset 135
Tabel 5. 36 Data token OOV pada KEPS 141
Tabel 5. 37 Data token OOV pada TermA 141