# DAFTAR ISI

**Halaman**

HALAMAN JUDUL i

HALAMAN PENGESAHAN ii

[SURAT PERNYATAAN KEASLIAN iii](#_Toc107180347)

[ABSTRAK iv](#_Toc107180348)

[ABSTRACT v](#_Toc107180349)

[KATA PENGANTAR vi](#_Toc107180350)

[DAFTAR ISI viii](#_Toc107180351)

[DAFTAR GAMBAR xi](#_Toc107180352)

[DAFTAR TABEL xiii](#_Toc107180353)

[DAFTAR SEGMEN PROGRAM xiv](#_Toc107180354)

[DAFTAR RUMUS xv](#_Toc107180355)

[BAB 1 PENDAHULUAN 1](#_Toc107180356)

[1.1 Latar Belakang 1](#_Toc107180357)

[1.2 Tujuan 3](#_Toc107180358)

[1.3 Batasan Penelitian 3](#_Toc107180359)

[1.4 Sistematika Pembahasan 4](#_Toc107180360)

[BAB 2 TEORI PENUNJANG 6](#_Toc107180361)

[2.1 Word Embedding 6](#_Toc107180362)

[2.2 Named Entity Recognition (NER) 8](#_Toc107180363)

[2.3 Transformer 9](#_Toc107180364)

[2.3.1 Bagian Encoder 10](#_Toc107180365)

[2.3.2 Bagian Decoder 14](#_Toc107180366)

[2.4 Neural Network (NN) 16](#_Toc107180367)

[2.4.1 Multi-Layer Perceptron (MLP) 16](#_Toc107180368)

[2.5 Sequence to Sequence Models 19](#_Toc107180369)

[2.6 Hungarian Match 20](#_Toc107180370)

[2.6.1 Simple Assignment Problem 21](#_Toc107180371)

[2.6.2 General Assignment Problem 25](#_Toc107180372)

[2.6.3 Hungarian Algorithm 28](#_Toc107180373)

[2.7 BRAT 30](#_Toc107180374)

[2.8 Macro dan Micro F1 Score 35](#_Toc107180375)

[BAB 3 NESTED NER DALAM BAHASA INDONESIA 38](#_Toc107180376)

[3.1 Named Enity Recognition (NER) 38](#_Toc107180377)

[3.2 Arsitektur Sistem 41](#_Toc107180378)

[3.3 Dataset dan Tagset Bahasa Inggris 49](#_Toc107180379)

[3.3.1 GENIA 51](#_Toc107180380)

[3.3.2 Penjelasan Jenis Tagset Bahasa Inggris 52](#_Toc107180381)

[3.4 Dataset dan Tagset Bahasa Indonesia 53](#_Toc107180382)

[3.4.1 Jenis Tagset Bahasa Indonesia 54](#_Toc107180383)

[3.4.2 Statistika Dataset 58](#_Toc107180384)

[3.5 Preprocessing 59](#_Toc107180385)

[3.5.1 Struktur dan Pelabelan Dataset 59](#_Toc107180386)

[3.5.2 Konversi dan Statistika Dataset 63](#_Toc107180387)

[BAB 4 SEQUENCE TO SET DALAM BAHASA INDONESIA 68](#_Toc107180388)

[4.1 Sequence To Set Network 68](#_Toc107180389)

[4.1.1 Sequence Encoder 70](#_Toc107180390)

[4.1.1.1 BERT 70](#_Toc107180391)

[4.1.1.2 Word2Vec 74](#_Toc107180392)

[4.1.1.3 Character-level BiLSTM 77](#_Toc107180393)

[4.1.1.4 BiLSTM 79](#_Toc107180394)

[4.1.1.1 Encoded Tokens 82](#_Toc107180395)

[4.1.2 Entity Set Decoder 83](#_Toc107180396)

[4.1.2.1 Decoder Layer 83](#_Toc107180397)

[4.1.2.2 Classification Layer 87](#_Toc107180398)

[4.1.3 Bipartite Matching 88](#_Toc107180399)

[4.2 Library dan Tools 88](#_Toc107180400)

[4.2.1 PyTorch 89](#_Toc107180401)

[4.2.2 Hugging Face (Transformers) 89](#_Toc107180402)

[4.2.3 Google Colab 90](#_Toc107180403)

[4.3 Modifikasi Metode Sequence To Set Network 91](#_Toc107180404)

[4.4 Contoh Kasus Penggunaan Sequence To Set 92](#_Toc107180405)

[BAB 5 UJI COBA 102](#_Toc107180406)

[5.1 Evaluator Performa Hasil dari Tugas Akhir 102](#_Toc107180407)

[5.2 Uji Coba Sequence-to-Set Network pada Nested NER 103](#_Toc107180408)

[5.2.1 Pengaruh Batch Size 104](#_Toc107180409)

[5.2.2 Pengaruh Learning Rate 105](#_Toc107180410)

[5.2.3 Pengaruh Gradient Norm 105](#_Toc107180411)

[5.2.4 Pengaruh Dropout 106](#_Toc107180412)

[5.3 Hasil Prediksi Sequence-to-Set Network 107](#_Toc107180413)

[5.4 Perbandingan Metode Span-Based 118](#_Toc107180414)

[5.5 Catatan Uji Coba 120](#_Toc107180415)

[BAB 6 PENUTUP 124](#_Toc107180416)

[6.1 Kesimpulan 124](#_Toc107180417)

[6.2 Saran 125](#_Toc107180418)

[DAFTAR PUSTAKA 126](#_Toc107180419)

[RIWAYAT HIDUP 130](#_Toc107180420)

[LAMPIRAN A A-1](#_Toc107180421)

[LAMPIRAN B B-1](#_Toc107180422)

# DAFTAR GAMBAR

Gambar Halaman

[2.1 Contoh Visualisasi (a) Bag of Words dan (b) Word Embedding 23](#_Toc104907033)

[2.2 Contoh Nilai Semantik Tersimpan dari Word Embedding 24](#_Toc104907034)

[2.3 Contoh Pengenalan Entitas Bernama 25](#_Toc104907035)

[2.4 Arsitektur Transformers 26](#_Toc104907036)

[2.5 Representasi Transformer bagian Encoder Decoder 27](#_Toc104907037)

[2.6 Contoh Isi dari Matriks Score 28](#_Toc104907038)

[2.7 Representasi Tiap Layer Dari Arsitektur Transformer 31](#_Toc104907039)

[2.8 Visualisasi Arsitektur MLP dengan Backpropagation 35](#_Toc104907040)

[2.9 Bentuk Ringkasan Arsitektur Seq2Seq 37](#_Toc104907041)

[2.10 Contoh Anotasi Text-Span pada BRAT 48](#_Toc104907042)

[2.11 Contoh Normalization Annotation 49](#_Toc104907043)

[3.1 Contoh Hasil Task Nested NER 56](#_Toc104907044)

[3.2 Arsitektur Sistem 57](#_Toc104907045)

[3.3 Arsitektur Sistem Pra Proses 58](#_Toc104907046)

[3.4 Tampilan File Teks pada BRAT 58](#_Toc104907047)

[3.5 Tampilan Window Fitur Pelabelan 59](#_Toc104907048)

[3.6 Struktur Data (a) Data Anotasi BRAT (b) Data Input Model 61](#_Toc104907049)

[3.7 Arsitektur Sistem Proses Prediksi Nested NER 62](#_Toc104907050)

[3.8 Arsitektur Sistem Pasca Proses 64](#_Toc104907051)

[3.9 Ontologi dan Statistika dari GENIA 67](#_Toc104907052)

[3.10 Contoh Struktur Dataset 77](#_Toc104907053)

[4.1 Arsitektur Sistem Sequence-To-Set Network 84](#_Toc104907054)

[4.2 Visualisasi Arsitektur BERT QnA 86](#_Toc104907055)

[4.3 Arsitektur Character-Level Embedding 93](#_Toc104907056)

[4.4 Arsitektur Recurrent Neural Network (RNN) 94](#_Toc104907057)

[4.5 (Satu) Memory Cell LSTM 96](#_Toc104907058)

[4.6 Arsitektur Bagian Entity Set Decoder 99](#_Toc104907059)

[4.7 Alur Kalkulasi Self-Attention 101](#_Toc104907060)

[5.1 Hasil Uji Coba Jumlah Batch Size 113](#_Toc104907061)

[5.2 Hasil Uji Coba Learning Rate 114](#_Toc104907062)

[5.3 Hasil Uji Coba Gradient Norm 114](#_Toc104907063)

[5.4 Hasil Uji Coba Dropout 115](#_Toc104907064)

# DAFTAR TABEL

Tabel Halaman

3.1 Spesifikasi Dataset GENIA 67

3.2 Daftar Tagset ACE 2005 68

3.3 Jenis Tagset Bahasa Indonesia 71

3.4 Spesifikasi Dataset NER Bahasa Indonesia 73

3.5 Statistika Dataset Tugas Akhir 82

5.1 Contoh Data Confusion Matriks 109

5.2 Penghitungan Micro dan Macro 110

5.3 Hasil Uji Coba F1 Score Terbaik 112

5.4 Tabel Perbandingan F1 Score 116

# DAFTAR SEGMEN PROGRAM

Segmen Program Halaman

[3.1 Preprocessing TXT File 78](#_Toc104907377)

[3.2 Preprocessing ANN File 80](#_Toc104907378)

[3.3 Preprocess Indeks ANN 81](#_Toc104907379)

# DAFTAR RUMUS

Rumus Halaman

[2.1 Rumus Softmax 29](#_Toc104907547)

[2.2 Rumus Attention 29](#_Toc104907548)

[2.3 Rumus Multi-Head Attention 29](#_Toc104907549)

[2.4 Rumus Residual Connection Genap 30](#_Toc104907550)

[2.5 Rumus Residual Connection Ganjil 30](#_Toc104907551)

[2.6 Rumus Fungsi Linear 33](#_Toc104907552)

[2.7 Rumus Fungsi Aktivasi 34](#_Toc104907553)

[2.8 Rumus Loss Function 36](#_Toc104907554)

[2.9 Rumus Nilai Gradien 36](#_Toc104907555)

[2.10 Rumus Tahap Backward Pass 36](#_Toc104907556)

[2.11 Rumus Adequate Budget 43](#_Toc104907557)

[2.12 Rumus Adequate Budget 44](#_Toc104907558)

[2.13 Rumus Adequate Budget (Sederhana) 44](#_Toc104907559)

[2.14 Rumus Teorema 6 44](#_Toc104907560)

[2.15 Rumus Corollary 1 44](#_Toc104907561)

[2.16 Rumus Corollary 1 (Sederhana) 45](#_Toc104907562)

[2.17 Rumus Corollary 1 (Sederhana 2) 45](#_Toc104907563)

[2.18 Rumus Perubahan budget 45](#_Toc104907564)

[4.1 Rumus Fungsi Y 95](#_Toc104907565)

[4.2 Rumus Fungsi H 95](#_Toc104907566)

[4.3 Rumus Fungsi F 95](#_Toc104907567)

[4.4 Rumus Fungsi G 95](#_Toc104907568)

[4.5 Rumus Gate Input, 96](#_Toc104907569)

[4.6 Rumus Gate Forget 96](#_Toc104907570)

[4.7 Rumus Vektor Sell 96](#_Toc104907571)

[4.8 Rumus Hidden Sell 96](#_Toc104907572)

[4.9 Rumus Token Awal 97](#_Toc104907573)

[4.10 Rumus Token Forward BiLSTM 98](#_Toc104907574)

[4.11 Rumus Token Backward BiLSTM 98](#_Toc104907575)

[4.12 Rumus Token 98](#_Toc104907576)

[4.13 Rumus Attention 102](#_Toc104907577)

[4.14 Rumus Multi-Head Attention 102](#_Toc104907578)

[4.15 Rumus Output Multi-Head Attention 102](#_Toc104907579)

[4.16 Rumus Probabilitas Class 103](#_Toc104907580)

[4.17 Rumus Probabilitas 103](#_Toc104907581)

[4.18 Rumus Probabilitas Boundary Kiri 103](#_Toc104907582)

[4.19 Rumus Probabilitas Boundary Kanan 103](#_Toc104907583)

[4.20 Rumus Indeks Optimal Matching 104](#_Toc104907584)

[4.21 Rumus Hungarian 104](#_Toc104907585)

[4.22 Rumus Fungsi Loss 104](#_Toc104907586)

[5.1 Rumus Precision 109](#_Toc104907587)

[5.2 Rumus Recall 109](#_Toc104907588)

[5.3 Rumus F1 Score 109](#_Toc104907589)