# DAFTAR ISI

**Halaman**

HALAMAN JUDUL i

HALAMAN PENGESAHAN ii

SURAT PERNYATAAN KEASLIAN iii

ABSTRAK iv

ABSTRACT v

KATA PENGANTAR vi

DAFTAR ISI viii

DAFTAR GAMBAR xi

DAFTAR TABEL xiii

DAFTAR SEGMEN PROGRAM xiv

DAFTAR RUMUS xv

BAB 1 PENDAHULUAN 1

1.1 Latar Belakang 1

1.2 Tujuan 3

1.3 Batasan Penelitian 3

1.4 Sistematika Pembahasan 4

BAB 2 TEORI PENUNJANG 6

2.1 Word Embedding 6

2.2 Named Entity Recognition (NER) 8

2.3 Transformer 9

2.3.1 Bagian Encoder 10

2.3.2 Bagian Decoder 14

2.4 Neural Network (NN) 16

2.4.1 Multi-Layer Perceptron (MLP) 16

2.5 Sequence to Sequence Models 19

2.6 Hungarian Match 20

2.6.1 Simple Assignment Problem 21

2.6.2 General Assignment Problem 25

2.6.3 Hungarian Algorithm 28

2.7 BRAT 30

2.8 Macro dan Micro F1 Score 35

BAB 3 NESTED NER DALAM BAHASA INDONESIA 38

3.1 Named Enity Recognition (NER) 38

3.2 Arsitektur Sistem 41

3.3 Dataset dan Tagset Bahasa Inggris 49

3.3.1 GENIA 51

3.3.2 Penjelasan Jenis Tagset Bahasa Inggris 52

3.4 Dataset dan Tagset Bahasa Indonesia 53

3.4.1 Jenis Tagset Bahasa Indonesia 54

3.4.2 Statistika Dataset 58

3.5 Preprocessing 59

3.5.1 Struktur dan Pelabelan Dataset 59

3.5.2 Konversi dan Statistika Dataset 63

BAB 4 SEQUENCE TO SET DALAM BAHASA INDONESIA 68

4.1 Sequence To Set Network 68

4.1.1 Sequence Encoder 70

4.1.1.1 BERT 70

4.1.1.2 Word2Vec 74

4.1.1.3 Character-level BiLSTM 77

4.1.1.4 BiLSTM 79

4.1.1.1 Encoded Tokens 82

4.1.2 Entity Set Decoder 83

4.1.2.1 Decoder Layer 83

4.1.2.2 Classification Layer 87

4.1.3 Bipartite Matching 88

4.2 Library dan Tools 88

4.2.1 PyTorch 89

4.2.2 Hugging Face (Transformers) 89

4.2.3 Google Colab 90

4.3 Modifikasi Metode Sequence To Set Network 91

4.4 Contoh Kasus Penggunaan Sequence To Set 92

BAB 5 UJI COBA 102

5.1 Evaluator Performa Hasil dari Tugas Akhir 102

5.2 Uji Coba Sequence-to-Set Network pada Nested NER 103

5.2.1 Pengaruh Batch Size 104

5.2.2 Pengaruh Learning Rate 105

5.2.3 Pengaruh Gradient Norm 105

5.2.4 Pengaruh Dropout 106

5.3 Hasil Prediksi Sequence-to-Set Network 107

5.4 Perbandingan Metode Span-Based 118

5.5 Catatan Uji Coba 120

BAB 6 PENUTUP 124

6.1 Kesimpulan 124

6.2 Saran 125

DAFTAR PUSTAKA 126

RIWAYAT HIDUP 130

LAMPIRAN A DATASET A-1

LAMPIRAN B ENVIRONMENT B-1

# DAFTAR GAMBAR

Gambar Halaman

2.1 Contoh Visualisasi Bag of Words dan Word Embedding 7

2.2 Contoh Nilai Semantik Tersimpan dari Word Embedding 7

2.3 Contoh Pengenalan Entitas Bernama 8

2.4 Arsitektur Transformers 9

2.5 Representasi Transformer bagian Encoder Decoder 10

2.6 Contoh Isi dari Matriks Score 11

2.7 Representasi Tiap Layer Dari Arsitektur Transformer 13

2.8 Visualisasi Arsitektur MLP dengan Backpropagation 17

2.9 Bentuk Ringkasan Arsitektur Seq2Seq 19

2.10 Contoh Anotasi Text-Span pada BRAT 31

2.11 Contoh Normalization Annotation 31

3.1 Contoh Hasil Task Nested NER 40

3.2 Arsitektur Sistem 42

3.3 Arsitektur Sistem Pra Proses 43

3.4 Tampilan File Teks pada BRAT 43

3.5 Tampilan Window Fitur Pelabelan 44

3.6 Struktur Data (a) Data Anotasi BRAT (b) Data Input Model 45

3.7 Arsitektur Sistem Proses Prediksi Nested NER 46

3.8 Arsitektur Sistem Pasca Proses 49

3.9 Ontologi dan Statistika dari GENIA 51

3.10 Contoh Struktur Dataset 62

4.1 Arsitektur Sistem Sequence-To-Set Network 69

4.2 Visualisasi Arsitektur BERT QnA 71

4.3 Arsitektur Character-Level Embedding 78

4.4 Arsitektur Recurrent Neural Network (RNN) 79

4.5 (Satu) Memory Cell LSTM 81

4.6 Arsitektur Bagian Entity Set Decoder 84

4.7 Alur Kalkulasi Self-Attention 85

5.1 Hasil Uji Coba Jumlah Batch Size 104

5.2 Hasil Uji Coba Learning Rate 105

5.3 Hasil Uji Coba Gradient Norm 106

5.4 Hasil Uji Coba Dropout 106

5.5 Hasil Prediksi Domain Politik 1 107

5.6 Hasil Prediksi Domain Politik 2 108

5.7 Hasil Prediksi Domain Politik 3 109

5.8 Hasil Prediksi Domain Hukum dan Kriminal 1 110

5.9 Hasil Prediksi Domain Hukum dan Kriminal 2 111

5.10 Hasil Prediksi Domain Hukum dan Kriminal 3 112

5.11 Hasil Prediksi Domain Peristiwa 1 112

5.12 Hasil Prediksi Domain Peristiwa 2 113

5.13 Hasil Prediksi Domain Peristiwa 3 114

5.14 Hasil Prediksi Contoh Kasus 1.1 114

5.15 Hasil Prediksi Contoh Kasus 1.2 115

5.16 Hasil Prediksi Contoh Kasus 2.1 115

5.17 Hasil Prediksi Contoh Kasus 2.2 116

5.18 Hasil Prediksi Contoh Kasus 3.1 116

5.19 Hasil Prediksi Contoh Kasus 3.2 117

5.20 Hasil Prediksi Contoh Kasus 4 118

# DAFTAR TABEL

Tabel Halaman

2.1 Contoh Data Confusion Matriks 35

2.2 Penghitungan Micro dan Macro 36

3.1 Spesifikasi Dataset GENIA 52

3.2 Daftar Tagset ACE 2005 53

3.3 Jenis Tagset Bahasa Indonesia 55

3.4 Spesifikasi Dataset NER Bahasa Indonesia 58

3.5 Statistika Dataset Tugas Akhir 67

5.1 Hasil Uji Coba F1 Score Terbaik 103

5.2 Tabel Perbandingan F1 Score dengan GENIA 119

5.3 Tabel Perbandingan F1 Score dengan Dataset Tugas Akhir 120

# DAFTAR SEGMEN PROGRAM

Segmen Program Halaman

3.1 Preprocessing TXT File 64

3.2 Preprocessing ANN File 65

3.3 Preprocess Indeks ANN 66

# DAFTAR RUMUS

Rumus Halaman

2.1 Rumus Softmax 12

2.2 Rumus Attention 12

2.3 Rumus Multi-Head Attention 12

2.4 Rumus Residual Connection Genap 13

2.5 Rumus Residual Connection Ganjil 13

2.6 Rumus Fungsi Linear 16

2.7 Rumus Fungsi Aktivasi 16

2.8 Rumus Loss Function 18

2.9 Rumus Nilai Gradien 18

2.10 Rumus Tahap Backward Pass 18

2.11 Rumus Adequate Budget 25

2.12 Rumus Adequate Budget 26

2.13 Rumus Adequate Budget (Sederhana) 26

2.14 Rumus Teorema 6 26

2.15 Rumus Corollary 1 27

2.16 Rumus Corollary 1 (Sederhana) 27

2.17 Rumus Corollary 1 (Sederhana 2) 27

2.18 Rumus Perubahan Budget 28

2.19 Rumus Precision 36

2.20 Rumus Recall 36

2.21 Rumus F1 Score 36

4.1 Rumus Fungsi Y 80

4.2 Rumus Fungsi H 80

4.3 Rumus Fungsi F 80

4.4 Rumus Fungsi G 80

4.5 Rumus Gate Input 81

4.6 Rumus Gate Forget 81

4.7 Rumus Vektor Sell 81

4.8 Rumus Hidden Sell 81

4.9 Rumus Token Awal 82

4.10 Rumus Token Forward BiLSTM 82

4.11 Rumus Token Backward BiLSTM 82

4.12 Rumus Token 83

4.13 Rumus Attention 87

4.14 Rumus Multi-Head Attention 87

4.15 Rumus Output Multi-Head Attention 87

4.16 Rumus Probabilitas Class 87

4.17 Rumus Probabilitas 87

4.18 Rumus Probabilitas Boundary 87

4.19 Rumus Probabilitas Boundary Kanan 87

4.20 Rumus Indeks Optimal Matching 88

4.21 Rumus Hungarian 88

4.22 Rumus Fungsi Loss 88