

HALAMAN PENGESAHAN	i
PERNYATAAN BEBAS PLAGIASI	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	v
DAFTAR ISI	vi
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR SINGKATAN.....	xiii
INTISARI.....	xiv
ABSTRACT	xv
BAB I Pendahuluan	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Batasan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan.....	4
BAB II Tinjauan Pustaka.....	6
2.1 Menjaga Privasi dengan <i>Data Masking</i>	6
2.2 Pembuatan Sistem NER Berbasis <i>Non-Transformer</i>	7
2.2.1 Pembuatan Model NER dengan <i>Encoder BiLSTM-CRF</i>	7
2.2.2 Pembuatan Model NER dengan <i>Encoder CNN-BiGRU</i>	8
2.2.3 Pembuatan Model NER dengan <i>Multi Hash Embedding</i> dan <i>Encoder WindowBasedEncoder</i>	10
2.2.4 Pembuatan Model NER dengan <i>Hash Embedding</i>	11
2.2.5 Pembuatan Model NER dengan <i>Character Embedding</i> dan <i>Encoder BiLSTM</i>	13
2.3 Pembuatan Sistem NER Berbasis <i>Transformer</i>	13
2.3.1 Pembuatan Model NER dengan BERT	13
2.4 Aplikasi <i>Data Masking</i> dengan Sistem NER pada PII	16
BAB III Dasar Teori	17
3.1 Data	17
3.2 Pengolahan Data	17
3.2.1 <i>Data Mining</i>	17
3.2.2 <i>Data Privacy</i>	18
3.2.3 <i>Data Masking</i>	18

3.2.4	<i>Data Redaction</i>	18
3.2.5	<i>Data Preprocessing</i>	19
3.2.6	<i>Data Annotation</i>	20
3.3	<i>Personally Identifiable Information</i>	20
3.4	Pemrosesan Bahasa Alami	21
3.5	Pembelajaran Mesin	21
3.6	<i>Named Entity Recognition (NER)</i>	22
3.7	Uji Evaluasi Model NER	23
3.8	<i>Deep Learning</i>	25
3.9	<i>Transformer</i>	25
3.10	Model <i>Deep Learning</i> Berbasis <i>Transformer</i>	28
3.10.1	<i>Bidirectional Encoder Representations from Transformers (BERT)</i>	28
3.11	Model <i>Deep Learning</i> Berbasis Non- <i>Transformer</i>	30
3.11.1	<i>Feed Forward Network</i>	30
3.11.2	<i>Convolutional Neural Network (CNN)</i>	32
3.12	Metode <i>Embedding</i>	33
3.12.1	<i>Hash Embedding</i>	33
3.12.2	<i>Multi Hash Embedding</i>	34
3.12.3	<i>Character Embedding</i>	34
3.13	Metode <i>Encoding</i>	35
3.13.1	<i>Bidirectional Long Short-Term Memory (BiLSTM)</i>	36
3.13.2	<i>Window-Based Encoder</i>	37
3.14	Fungsi Aktivasi	38
3.14.1	<i>Maxout Activation Function</i>	38
3.14.2	<i>Mish Activation Function</i>	39
3.15	MHE-WindowBasedEncoder(Maxout)	40
3.16	HE-WindowBasedEncoder(Maxout)	40
3.17	CE-WindowBasedEncoder(Maxout)	41
3.18	CE-WindowBasedEncoder(Mish)	41
3.19	CE-BiLSTM(Softmax)	42
3.20	<i>Transformer Huggingface-IndoBERT</i>	42
BAB IV	Metode Penelitian	44
4.1	Alat dan Bahan Tugas akhir	44
4.1.1	Alat Tugas akhir	44
4.1.2	Bahan Tugas akhir	45
4.2	Metode yang Digunakan	45
4.3	Alur Tugas Akhir	46
4.4	Pembuatan Dataset NER	49
4.4.1	Penentuan Sumber Dataset	49

4.4.2	Pembuatan <i>Framework Web Scraping</i>	49
4.4.3	Verifikasi Dataset.....	52
4.4.4	<i>Preprocessing</i> Data	53
4.4.5	Pengolahan Entitas Sensitif Dataset.....	55
4.4.5.1	Pembuatan Dataset Informasi Sensitif	55
4.4.5.2	Pengintegrasian ke dalam Dataset Asli	56
4.4.6	Pelabelan Dataset	56
4.5	Pembuatan Model NER.....	58
4.5.1	Penentuan Versi SpaCy	58
4.5.2	Pembuatan <i>Pipeline</i> SpaCy	58
4.5.3	Pembuatan 7 Model NER.....	60
4.5.4	Pembuatan Metode <i>Masking</i>	61
4.6	Pengujian Model NER.....	61
4.6.1	Pengujian Model NER dengan Skema Label BILOU	62
BAB V	Hasil dan Pembahasan.....	64
5.1	Hasil Entitas Dataset.....	64
5.2	Hasil Evaluasi Model <i>Deep Learning</i> NER Berbasis Non- <i>Transformer</i>	64
5.2.1	MHE-WindowBasedEncoder(Maxout)	64
5.2.2	HE-WindowBasedEncoder(Maxout).....	66
5.2.3	CE-WindowBasedEncoder(Maxout)	67
5.2.4	CE-BiLSTM(Softmax, <i>drop</i> =0)	69
5.2.5	CE-BiLSTM (Softmax, <i>drop</i> =0.2)	70
5.2.6	CE-WindowBasedEncoder(Mish).....	72
5.3	Hasil Evaluasi Model <i>Deep Learning</i> NER Berbasis <i>Transformer</i>	73
5.3.1	<i>Transformer</i> Huggingface-IndoBERT	73
5.4	Perbandingan Hasil Evaluasi Model NER	75
5.4.1	Perbandingan Akurasi	75
5.4.2	Perbandingan Hasil Implementasi.....	77
5.4.2.1	Implementasi Kasus 1	77
5.4.2.2	Implementasi Kasus 2	81
5.5	Perbandingan Kinerja Berdasarkan Hasil Implementasi	86
5.6	Hasil Implementasi <i>Masking</i> pada Teks Bahasa Indonesia	87
BAB VI	Kesimpulan dan Saran.....	89
6.1	Kesimpulan.....	89
6.2	Saran.....	90
DAFTAR PUSTAKA.....		91
LAMPIRAN		L-1
L.1	Kode Sumber	L-1
L.1.1	Kode Sumber <i>Web Scraping</i>	L-1

L.1.2	Kode Sumber <i>Data Preprocessing</i>	L-2
L.1.3	Kode Sumber Utama	L-4