

DAFTAR ISI

Halaman Judul	ii
Halaman Pengesahan	iii
Halaman Pernyataan	iv
Halaman Persembahan	v
Halaman Motto	vi
PRAKATA	vii
INTISARI	xvii
ABSTRACT	xviii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan masalah	3
1.3 Batasan masalah	3
1.4 Tujuan penelitian	4
1.5 Manfaat penelitian	4
1.6 Metodologi Penelitian	4
1.7 Sistematika Penulisan	5
II TINJAUAN PUSTAKA	7
III LANDASAN TEORI	11
3.1 <i>Voice Activity Detection</i>	11
3.2 <i>Forced Alignment</i> pada Sistem <i>Automatic Speech Recognition</i>	13
3.3 Ekstraksi Fitur	14
3.3.1 <i>Log-Mel Filterbank</i> (log-mel)	15
3.3.2 Delta	16
3.4 Jaringan Syaraf Tiruan	16
3.4.1 Bias	18

3.4.2	Fungsi Aktivasi	18
3.4.3	Fungsi Biaya	19
3.4.4	Gradient Descent	20
3.5	RMSProp	21
3.6	<i>Recurrent Neural Network</i>	22
3.6.1	<i>Long Short-Term Memory</i>	25
3.6.2	<i>Gated Recurrent Unit</i>	27
3.7	Dropout	29
3.8	<i>Receiver Operating Characteristic (ROC) dan Area Under ROC Curve (AUC)</i>	30
IV ANALISIS DAN PERANCANGAN		33
4.1	Analisis Permasalahan	33
4.2	Rancangan Penelitian	33
4.3	Rancangan Data	35
4.3.1	Pelabelan VAD	36
4.3.2	Sintesis suara pengucapan dengan noise	40
4.4	Rancangan Pra-pemrosesan	43
4.4.1	Ekstraksi fitur	43
4.4.2	Pemrosesan label	44
4.4.3	Pembentukan <i>mini-batch</i> dan <i>padding</i>	45
4.5	Rancangan Arsitektur	46
4.5.1	<i>Input layer</i>	47
4.5.2	<i>Hidden layer</i>	47
4.5.3	<i>Output layer</i>	47
4.6	Rancangan Pelatihan	48
4.6.1	<i>Learning rate</i> dan <i>decay rate</i>	48
4.6.2	Dropout <i>keep probability</i>	49
4.6.3	<i>Mini-batch, padding, dan masking loss</i>	49
4.6.4	<i>Truncated step</i>	49
4.6.5	Jumlah <i>epoch</i> dan <i>stop condition</i>	50
4.6.6	Inisialisasi parameter bobot dan bias	50
4.7	Rancangan Eksperimen dan Pengujian	50
4.7.1	Eksperimen pencarian <i>hyper-parameter</i> terbaik	51
4.7.2	Pengujian terhadap tingkatan <i>noise</i> dan arsitektur lain	53

V	IMPLEMENTASI	54
5.1	Spesifikasi Perangkat Keras dan Lunak	54
5.2	Pelabelan Data	55
5.2.1	Pelatihan sistem ASR menggunakan Kaldi	55
5.2.2	<i>Forced Alignment</i> menggunakan sistem ASR	57
5.2.3	Modifikasi label menjadi <i>speech</i> dan <i>non-speech</i>	57
5.3	Sintesis Suara Pengucapan	58
5.4	Pra-pemrosesan <i>Dataset</i>	63
5.4.1	Iterator pra-pemrosesan data	67
5.4.2	Ekstraksi fitur	69
5.4.3	Pemrosesan label	70
5.5	Implementasi Menggunakan Tensorflow	71
5.5.1	Arsitektur RNN	72
5.5.2	Arsitektur GRU, LSTM, dan RNN dengan aktivasi tanh	73
5.5.3	Optimisasi fungsi biaya	75
5.5.4	Pelatihan dan prediksi secara <i>truncated</i>	77
5.5.5	Metrik evaluasi AUC	78
VI	HASIL DAN PEMBAHASAN	83
6.1	Proses Voice Activity Detection	83
6.2	Eksperimen Pencarian <i>Hyper-parameter</i> Terbaik	84
6.2.1	Eksperimen pemilihan konfigurasi mula-mula	84
6.2.2	Eksperimen pemilihan <i>learning rate</i> dan <i>truncated step</i>	87
6.2.3	Perbandingan <i>hyper-parameter</i> terbaik dari setiap arsitektur	97
6.3	Pengujian Terhadap Tingkatan Noise	103
VII	KESIMPULAN DAN SARAN	110
7.1	Kesimpulan	110
7.2	Saran	110
	DAFTAR PUSTAKA	112